

# THIS WEEK

## EDITORIALS

**DRUGS** Too many policies are based on little more than prejudice **p.166**

**WORLD VIEW** Assessment reports need a fresh approach **p.167**



**VALHALLA** Vikings took headless slaves to the afterlife **p.168**

## Precautionary measures

*Major African campaigns targeting malaria and HIV could help millions, but key concerns over their long-term effects should not be forgotten.*

More than one million healthy children in Africa participated in a bold initiative this year when they took antimalarial drugs during the rainy season. The hope is that a few months of preemptive doses could help them to fight off a disease that kills some 600,000 people throughout Africa each year. At the same time, health campaigns in 14 nations in eastern and southern Africa are circumcising millions of men in an attempt to stem the spread of HIV, which infects more than one-quarter of people aged 15–49 in some of those countries. The malaria and HIV efforts have the potential to gain the upper hand on two stubborn health scourges, but only if funders and organizers heed the lessons of past failures.

*Nature* this week publishes two reports from the ground. They are written by reporters who travelled to Africa to examine the benefits of these health campaigns — as well as concerns that have emerged about them. In the malaria effort (see page 186), one of the biggest worries has been that giving medications to children will promote the spread of resistant forms of the malaria parasite, quickly rendering the drugs ineffective. That happened during the 1950s and 1960s, when doctors performed prevention experiments for malaria in Africa and South America. The current campaign is based on trials that started in 2002 and took steps to avoid spreading resistance to front-line treatments by providing a cocktail of older antimalarial drugs, and only during the rainy season.

The success seen in the clinical trials, however, is not guaranteed as the programme is scaled up to cover possibly more than 20 million children in parts of Africa. Six nations started giving antimalarials this year, but treated just a fraction of their intended recipients because of funding and organization problems. This raises concerns that the programmes will not carry out the ancillary monitoring efforts needed to ensure success. Funding must be provided to track whether the large-scale prevention campaign reduces the number of malaria cases as hoped, and to ensure that resistant forms of the parasite do not spread more quickly than anticipated. The problem is that funders are typically less interested in supporting follow-up studies than in testing ideas and carrying out interventions. And monitoring has long been a weak point for malaria: global surveillance catches just 10% of the estimated global malaria cases each year.

In the circumcision campaign (see page 182), several trials showed that the procedure reduces the risk of HIV transmission from women to men by a substantial 50–60%. That level of protection is so large that international aid organizations, the United Nations and donor countries such as the United States have poured more than US\$100 million into campaigns that seek to circumcise 20 million men in 14 target countries by 2015. But some researchers worry that the massive programmes will not yield the same benefits as the small, intensive trials. A prime concern is the message that men and women are getting about the effectiveness of the procedure. Campaign organizers have mounted major advertising efforts to encourage more acceptance

of circumcision. In Zambia, billboards proclaim that a circumcised individual is ‘a man who cares’. In Tanzania, he is ‘60% more man’.

The advertisements have succeeded in getting millions of men through the clinic doors, but the messages have also generated confusion. Many women in several of the target countries assume incorrectly that circumcision helps to protect them against acquiring HIV, and so they think that condoms are less necessary than they once were. Several studies have shown that men overestimate the amount of protection they gain through circumcision. And in Tanzania, the phrase ‘60% more man’ is taken to mean that a circumcised man has more sexual partners — not the kind of message that will cut down on HIV transmission.

During the preliminary trials, clinics took steps to avert confusion by providing substantial counselling about the risks and benefits of circumcision before and after the procedure and in periodic follow-up visits. Men received testing and treatment for HIV and for several other sexually transmitted infections. But in the scaled-up campaign, circumcision providers typically give only one counselling session related to the procedure, and another with an HIV test. Behavioural researchers say that men require more counselling and that campaign organizers should also target information towards women to clear up misconceptions.

Just as with the preventive malaria drugs, there must be sufficient monitoring to track whether circumcision is as effective at reducing HIV transmission as it was during the smaller trials. Early signs are positive, and that is good news for millions of African men and women. ■

**“Funders are typically less interested in supporting follow-up studies.”**

## Data deadline

*Time is running out to comment on the NIH’s plan for sharing genomic data.*

A little-noticed proposal promises to have a huge impact on how science is done in the ‘big data’ era. In September, the US National Institutes of Health (NIH) released draft guidelines on the sharing of genomic data. The guidelines, which have been in the works for five years, are a necessary and valuable update to the agency’s stance on how researchers who receive its funds must share data produced by projects that use array-based and high-throughput technologies. They cover a huge swathe of research, including sequencing human and non-human genomes, genes and gene variants, as well as transcriptomic, epigenomic and gene-expression data.

The issues related to collecting and sharing such data are complex,

and the guidelines touch on most of the controversial topics in large-scale biology research today.

One of the key issues is when to share. The draft policy says that researchers must have shared their data by the time those data are published in a formal manuscript. However, there are earlier release deadlines for some data types, such as raw sequence data from non-human organisms and the initial analysis of some human sequence data, both of which must be shared within six months of submission to an approved repository. This is a thorny issue, and the NIH cannot please everybody. Some researchers favour the early release of more data, whereas others fear that releasing data ahead of publication will leave them vulnerable to being scooped.

Another major issue is how to protect the identity of those whose data are shared — especially as it is now clear that it is possible to identify people from anonymous data (see *Nature* <http://doi.org/px4>; 2013). The guidelines say that researchers should tell study participants that their data “may be shared broadly for future research purposes”, and let them know whether it will be shared through an open- or controlled-access mechanism. It asks researchers to gain explicit consent from patients who agree to share their data through open-access mechanisms. And, importantly, it sets a new bar for informed consent on de-identified materials, including cell lines and clinical specimens. Such research has historically been exempt from informed-consent requirements, but the guidelines ask researchers to obtain consent for future research on these materials, too.

This is a potentially major step, and one that this publication supports (see *Nature* 486, 293; 2012). It is true that some researchers who have relied on clinical specimens will see it as an impediment to valuable research. But similarly, some advocates of more transparent informed-consent rules will not like the fact that the guidelines allow researchers to opt out of this requirement if they give “compelling scientific reasons” for so doing.

A third aspect relates to how long the data should be shared for. Researchers who rely on controlled-access data sets often complain about periodically having to renew their requests for access. The guidelines maintain this standard, offering access to such data for one year at a time. This is unlikely to please those who have argued that legitimate scientists should be able to access larger tranches of data and for longer periods of time — although the NIH has responded that scientists who take this position are sometimes not aware of the restrictions on all of the data sets that they plan to use (see *Nature* 497, 172–174; 2013).

**“This is a thorny issue, and the NIH cannot please everybody.”**

Once finalized, the regulations will become part of a patchwork of international research regulation on the sharing of genomic data.

The United Kingdom, for instance, is still deciding how much information from its 100K Genome Project will be released and whether researchers will be able to access both sequencing results and the relevant personal health records. At the same time, US open-genomics evangelist George Church is expanding his Personal Genome Project to Canada and Europe, raising questions such as whether the project will be able to access records from centralized health systems (see [go.nature.com/izmgpo](http://go.nature.com/izmgpo)). By contrast, informed-consent regulations in other parts of the world are still being developed, leaving a question mark over whether the United States will become an easier place for genomics researchers to work than other parts of the world.

In that context, the US proposals will have a major impact on the work of *Nature's* readers. Yet, according to the NIH's Office of Science Policy, as of 7 November, just 18 comments had been received on the guidelines. That is a poor response to such an important issue. The policy will affect many more scientists and *Nature* urges them to submit their responses to the proposals before the deadline of 20 November. ■

# Keep asking

*Prejudice, not evidence, is too often the basis for government drug policies.*

Rob Ford, mayor of Toronto, Canada, caused a sensation last week when he told journalists, “Yes, I have smoked crack cocaine” — and refused to resign. The reporters smelled blood: Ford had long denied drug use, despite repeated rumours. “I wasn't lying,” he said when confronted. “You didn't ask the correct questions.”

The debate over the control and regulation of drugs is typified, perhaps more than any other in science policy, by a need to ask the correct questions. Politicians and the moralizing media tend to seek the black-or-white certainty of whether or not a drug poses a threat. Researchers often prefer to present risks as relative, and some argue that it is hypocritical to proscribe one compound while promoting, however tacitly, the consumption of another that may have similar — or more potent — effects.

From time to time the answer changes because of other factors, and this is where the question becomes less important than who is asking it. The television drama *Breaking Bad* has brought the abuse of methamphetamine to popular attention, but the government of Japan, for example, has long been well aware of the drug's effects. In fact, during the Second World War it encouraged the manufacture of the compound and distributed it to the country's soldiers and civilian workers in a handy tablet form called *hiropo*n. “For night work and other times demanding mental alertness. For overexertion,” a typical wartime advert read. “The most powerful new amphetamine on the market!”

At the end of the war, Japanese manufacturers sold the stimulant as

a cure for all manner of civilian ills, targeting, among others, juveniles disturbed by the country's dramatic post-war social change. Yet within a few years, a government U-turn introduced tough laws making the drug illegal, with harsh penalties for possession. Official propaganda now called on citizens to help the authorities to “wipe out the evil of stimulant drugs!”. Widespread abuse and signs of addiction gave the authorities a legitimate reason to act. But, as historian Jeffrey Alexander of the University of Wisconsin–Parkside pointed out in a paper this year (J. W. Alexander *Int. J. Drug Policy* 24, 238–243; 2013), there was another, more sinister motive: the deliberate cultivation of a media-fuelled drug panic to justify the arrest and deportation of Korean and Taiwanese immigrants, who were disproportionately blamed for making and selling methamphetamine.

Similar social pressures played a part in the crafting of US legislation against marijuana, which was first popular with Mexican labourers and black musicians. Prejudice is one of a number of issues contributing to policies on drugs that are explored by two books reviewed by Andrew Robinson on page 194. The books explain, he says, that “a drug's acceptability to mainstream society fluctuates more owing to social and cultural trends than to medical knowledge”. For example, the United States infamously banned alcohol during the prohibition era of the 1920s and 1930s, a move that would have been unlikely in the United Kingdom because of the “complex British attitude to drunkenness”.

This attitude was highlighted in 2009 by the UK government's then-drug adviser David Nutt, who argued publicly that alcohol and tobacco were more harmful than LSD, ecstasy and cannabis. His opinion earned him the sack. Last week, it also earned him the 2013

John Maddox Prize for Standing Up for Science, which recognizes the promotion of science in the public interest, and was set up with the help of this publication. Someone has to keep asking the correct questions. ■

➔ **NATURE.COM**  
To comment online,  
click on Editorials at:  
[go.nature.com/xhunjv](http://go.nature.com/xhunjv)



NICOLA CROCKFORD



## Review by quality not quantity for better policy

Global assessments need to adopt more rigorous and focused processes for collation and review, says William J. Sutherland.

Society increasingly demands that policies are based on scientific evidence, as it should. Yet policy-makers rarely refer to journal papers — the actual evidence — instead preferring to base decisions on summaries, reviews and assessments.

Such assessments can be deeply influential but fallible: the Intergovernmental Panel on Climate Change (IPCC) has effectively underpinned climate-change policies worldwide, yet the infamous Himalayan glacier blunder in the 2007 report is still regularly unearthed to taunt the IPCC's authority.

Less likely to make international headlines, but still instructive, are problems my colleague Lynn Dicks and I discovered when we reviewed the pollinator section of the United Kingdom's National Ecosystem Assessment, a similarly sprawling and influential report that fed into the British government's white paper on the environment. It states, for example, that "since 1980, wild bee diversity has declined in most landscapes". This exaggerates the findings of the cited paper, which reports only that bee diversity is lower post-1980 than pre-1980 in 52% of 81 10 × 10 kilometre grid squares. The report asserts that "decreases in pollination services would, therefore, result in short-term economic losses for farmers", but it fails to mention that this is based on indirect field studies of insect visitors to crops. In fact, no economic effects of pollinator decline have been detected on a national or sub-national scale. A crucial table in the assessment report that estimates the value of insect-pollinated crops in the United Kingdom comes from a PhD thesis, and the analysis behind it has not yet been published in the peer-reviewed scientific literature; this is not made clear. The same quality-control issues appear elsewhere in the report and in the Millennium Ecosystem Assessment.

The problem of quality control will continue unless we change the way in which such assessments are carried out. A good place to start is with the new Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES), which will meet early next month in Turkey to establish its assessment programme.

In broad terms, such assessments must be less ambitious. The headline conclusions of these reports are subjected to extensive scrutiny, yet they are linked to a vast expanse of text that is impossible to check thoroughly. The greater the intended scope of the assessment, and the more material that needs to be included for the sake of completeness, the higher the risk of error becomes. The pleading from climate scientists that the glacier mistake was just one paragraph among 938 pages in one of four reports, and was not even noticed for two years, has not quietened critics of the IPCC.

Although there might be a desire among some involved in the forthcoming IPBES assessment to write freely about many aspects of pollination,

they should instead concentrate on collating and synthesizing the global evidence on just the key issues relating to the main questions. Which pollinators are most economically important? What have been the changes to their status? What are the main threats? And what are the most economically effective means of maintaining and restoring these pollinators? To support the answers to these questions, the assessment must first produce a systematic and publicly available review of the literature it will be based on, ideally with accessible summaries of papers.

For such reviews, the vision is often the meta-analyses of evidence-based medicine, which pool and offer an overview of all studies of appropriate quality. This approach is less suited to environmental reviews, which typically use a complex mix of experiments, correlative studies and models with a range of questions, response variables and locations that have differing relevance to the issue of interest.

To better synthesize the findings of such diverse research, environmental assessments should use formal consensus methods such as the Delphi technique, which is an iterative process of consensus-seeking that is based on sequential rounds of confidential scoring and commenting on anonymized results. In addition, all of the material should be publicly available — possibly even including the individual final scores of the experts, as happens for the members of the UK Monetary Policy Committee who meet monthly to decide the official

interest rate. It would thus be relatively straightforward to trace a conclusion to its assessment of the evidence and then to its evidence base.

This approach seems to have a range of advantages. It concentrates on the few crucial issues but presents them in a more transparent and rigorous manner that is likely to provide greater confidence and reduce the likelihood of errors. After the evidence is collated, it can be updated regularly to allow for quick reassessment (conversely, the IPCC assessment is repeated about every six years and is hugely expensive). With the bedrock of the evidence assembled and presented in a user-friendly way, the evidence on key issues can be continually collated and regularly assessed. An equivalent example in medicine is *Clinical Evidence*, which reassesses interventions every six months.

This approach should be adopted not just by the IPBES and other future national or international assessments, but, where possible, by the innumerable reports that are written for decision makers. This should increase quality, transparency and authority. ■

**William J. Sutherland** is the Miriam Rothschild Professor of Conservation Biology in the Department of Zoology, University of Cambridge, UK.  
e-mail: [w.sutherland@zoo.cam.ac.uk](mailto:w.sutherland@zoo.cam.ac.uk)

ENVIRONMENTAL  
ASSESSMENTS  
SHOULD USE  
FORMAL  
CONSENSUS  
METHODS.

➔ **NATURE.COM**  
Discuss this article  
online at:  
[go.nature.com/owgnqw](http://go.nature.com/owgnqw)

# RESEARCH HIGHLIGHTS

Selections from the  
scientific literature

## CLIMATE SCIENCES

### Economic link to global warming

Changes in the rate of global warming can be pinned to specific episodes in human history.

In a statistical analysis, a team led by Francisco Estrada of the National Autonomous University of Mexico in Mexico City looked at global mean temperatures since 1880, as well as greenhouse-gas levels and other factors that warm or cool the planet. The authors found that economic slowdowns during the two world wars and in the Great Depression temporarily slowed the planet's warming.

The analysis also suggests that the Montreal Protocol, which phased out chemicals that deplete the ozone layer and trap heat, has helped to slow warming in recent decades.

*Nature Geosci.* <http://doi.org/p2b> (2013)

For a longer story on this research, see [go.nature.com/g93gb6](http://go.nature.com/g93gb6)

## ARCHAEOLOGY

### Burials indicate Viking sacrifices

Vikings may have sacrificed slaves to be buried with their masters.

In the 1980s, archaeologists working on the Norwegian island of Flakstad discovered and excavated ten individuals from six burials dating to between AD 550 and AD 1030. Remains from three multiple burials all contained at least



one skeleton lacking its skull.

To establish the relationship between them, a team led by Elise Naumann of the University of Oslo analysed mitochondrial DNA (which is inherited from the mother) as well as nitrogen and carbon isotopes from various parts of the skeletons (including a tooth; **pictured**).

The results suggested that individuals in the multiple burials did not share maternal ancestry, and that individuals found without skulls generally ate less meat than intact individuals in the same graves. The differences in ancestry and diet imply that the beheaded individuals might have been slaves, as has been suggested for other double burials from the Norse world. *J. Archaeol. Sci.* 41, 533–540 (2013)



## PHYSICS

### Traffic jams follow the laws of physics

Traffic congestion closely resembles the physics of phase transitions, such as when ice melts or a metal becomes superconducting.

Shin-ichi Tadaki at Saga University in Japan and his colleagues used a high-resolution laser scanner to track cars travelling around an empty indoor baseball stadium, then analysed those

data as if they were studying phase transitions in a material. They found that above a critical density of cars, traffic flow became unstable and changed from free-flowing to a jam.

Scaled up, that density value fits with those seen on real-world motorways, the authors say. *New J. Phys.* 15, 103034 (2013)

## CANCER

### Breast-cancer mutations found

Mutations in the gene encoding the oestrogen receptor (ER) have been identified in aggressive breast tumours.

About 70% of breast cancers express the ER, which helps to propel rampant growth. Several drugs target the effects of the sex hormone on its receptor, but tumours often develop resistance to such treatment.

Independent teams led by Sarat Chandralapathy at the Memorial Sloan-Kettering Cancer Center in New York, Arul Chinnaiyan at the University of Michigan in Ann Arbor and Ido Wolf at the Tel Aviv Sourasky Medical Center

in Israel, sequenced tumours from a total of 104 patients with metastatic breast cancer and found ER mutations in 25. Further work indicated that these mutations could keep the receptor active even in the absence of oestrogen.

Drugs that inhibit the receptor in other ways might stall aggressive tumours, the researchers say. *Nature Genet.* <http://doi.org/pxp>; <http://doi.org/pxq> (2013) *Cancer Res.* <http://dx.doi.org/10.1158/0008-5472.can-13-1197> (2013)

## CRYPTOGRAPHY

### Keeping quantum secrets

Quantum cryptographers have used devices in Switzerland and Singapore to deliver a

AKIHIRO NAKAYAMA

ELISE NAUMANN

secret message in the form of a committed bit — the value of which is kept hidden until a designated time.

Bit commitment is potentially useful for voting and other applications in which privacy is key, but ultra-secure quantum commitment schemes have proved difficult to implement.

Felix Bussi eres of the University of Geneva in Switzerland and his colleagues experimentally demonstrated a protocol proposed in 2012 for keeping committed bits secret. The technique, which relies on the constancy of the speed of light, can prevent insecure communication between remote locations for periods of up to 15 milliseconds, perhaps long enough to enable high-speed stock trading.

*Phys. Rev. Lett.* 111, 180504 (2013)

#### REGENERATIVE MEDICINE

## Embryo protein has healing power

A protein that is present mainly in developing embryos gives some adult tissues an exceptional ability to heal.

The protein Lin28a controls growth and development in animals from worms to humans, but it is rarely found in adult tissues in vertebrates.

When researchers led by George Daley at Boston Children's Hospital in Massachusetts engineered mice so that the protein was produced into adulthood, the animals' hair grew faster and puncture wounds in their ears healed more thoroughly. Newborn animals were more likely to regrow the tips of severed digits.

Lin28a represses an anti-growth regulatory program and boosts the production of many metabolic enzymes. In doing so, the researchers say, the protein induces a state that enhances tissue repair. *Cell* 155, 778–792 (2013)  
For a longer story on this research, see [go.nature.com/8cejq3](http://go.nature.com/8cejq3)

#### INFLAMMATORY DISEASE

## Sun synchronizes immune system

Daily light cycles help to set animals' response to infection.

Lora Hooper of the University of Texas Southwestern Medical Center in Dallas and her colleagues showed that a protein that controls whether certain white blood cells mature is in turn controlled by proteins that set daily rhythms. When mice were exposed to abnormal light cycles, rhythm-setting proteins sent signals to the immune system, prompting it to make more inflammatory cells in the gut.

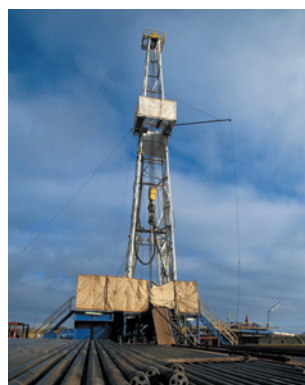
Inflammation protected the mice against bacterial infection, but it also made them more susceptible to inflammatory bowel disease and colitis. The findings might explain why shift workers and others with disrupted sleep cycles are more prone to immune disorders. *Science* 342, 727–730 (2013)

#### SEISMOLOGY

## Injected gas makes Earth rumble

Earthquakes that shook an old Texas oil field over the past few years could be due to injections of carbon dioxide gas. Previous tremors in the area were tied to wastewater injections, but these are the first events attributed to forcing pressurized gas underground.

Two years after significant gas injections began in the oil field to boost production



## COMMUNITY CHOICE

The most viewed papers in science

#### VIROLOGY

## New flu virus found in bats

**HIGHLY READ**  
on [plos.org](http://plos.org) in the past 30 days

The flat-faced fruit bat (*Artibeus planirostris*; pictured) in Peru carries a flu virus that was previously unknown to science.

H18N11 is the second influenza A virus to be identified in bats, and both were found by a team led by Ruben Donis of the Centers for Disease Control and Prevention in Atlanta, Georgia. The team found huge genetic diversity in bat influenza A viruses: four of eight bat viral gene segments harboured more genetic variation than did corresponding viral segments in other mammals and birds combined. Blood analyses also revealed that H18N11 infects several bat species.



Neither of the bat viruses has been found in humans, but the discoveries suggest that bats are a reservoir for diverse flu viruses that could fuel new infections. *PLoS Pathog.* 9, e1003657 (2013)

(pictured), the area experienced a spate of small earthquakes, 18 of which reached magnitudes between 3 and 4.4. Wei Gan and Cliff Frohlich at the University of Texas at Austin have now determined the precise location and timing of the quakes.

The locations correlate with gas injection sites and suggest that slippage occurred along a previously unidentified fault. Nearby fields where gas was also injected did not show such seismic activity, however, and the researchers say that more data are needed to predict where gas injection could trigger tremors.

*Proc. Natl Acad. Sci. USA*  
<http://doi.org/pxs> (2013)  
For a longer story on this research, see [go.nature.com/x4hnn4](http://go.nature.com/x4hnn4)

#### MEDICAL MICROBIOLOGY

## Gut microbes linked to arthritis

Unfriendly bacteria in the gut might bring on inflammation in the joints.

Dan Littman at the New York University School of

Medicine and his colleagues collected stool samples from people with arthritis. Sequencing experiments on gut bacteria showed that individuals with newly diagnosed arthritis were more likely to harbour the bacterium *Prevotella copri* than were individuals without the condition.

High levels of *P. copri* and related species correlated with low levels of beneficial microbes that are thought to suppress the immune system and metabolize vitamins into forms that are absorbed by the bloodstream. When inoculated with *P. copri*, mice became more sensitive to colitis, and the levels of beneficial microbes in their guts dropped.

Gut bacteria have long been implicated in arthritis, but this is the first time that researchers have correlated the onset of human disease with a particular microbial species. *eLife* 2, e01202 (2013)

**NATURE.COM**

For the latest research published by Nature visit:  
[www.nature.com/latestresearch](http://www.nature.com/latestresearch)



# SEVEN DAYS

The news in brief

## POLICY

### Shutdown scars

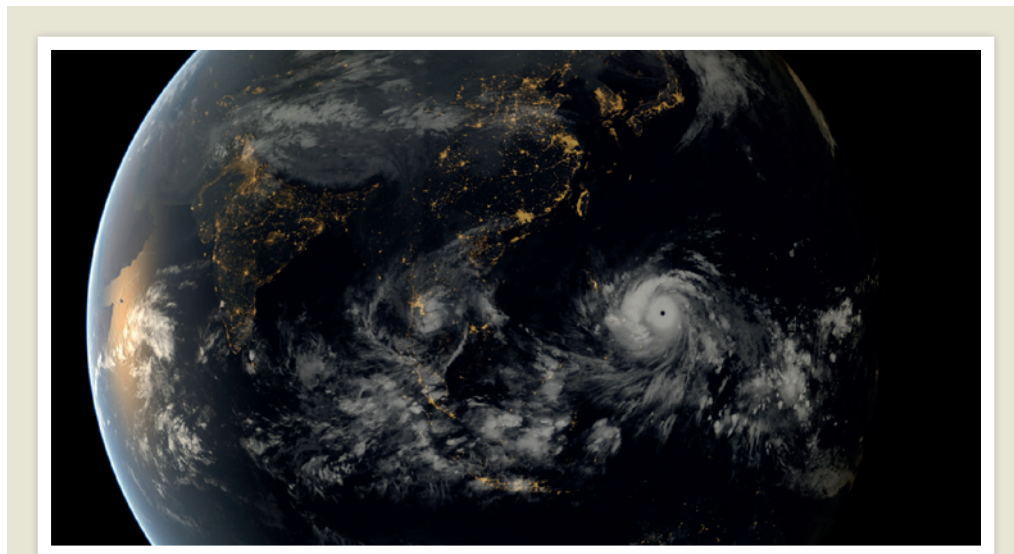
Last month's shutdown of the US government, during which many 'nonessential' federal workers were sent home, resulted in 6.6 million lost workdays and cost US\$2 billion in staff pay, the White House Office of Management and Budget said on 7 November. The 16-day shutdown, which began on 1 October, also curtailed the US Antarctic programme field season, cut back influenza monitoring and resulted in the loss of observing time at major telescope facilities. See [go.nature.com/u5kntz](http://go.nature.com/u5kntz) for more.

### Fat ban

On 7 November, the US Food and Drug Administration (FDA) proposed eliminating most artificial trans-fats — forms of fat associated with increased heart risk — from processed foods. The agency issued a preliminary determination that trans-fats in the form of partially hydrogenated oils should no longer be "generally recognized as safe", a status that allows a food additive to be used without FDA approval. A final determination will be made after a 60-day public comment period.

### Label loss

Voters in Washington state on 5 November rejected a ballot initiative to require labelling of genetically modified (GM) foods. The initiative was one of a string of efforts in more than 20 states this year to label or limit GM foods; California voters defeated a similar proposition last year. No state currently requires labelling of all GM foods, although Alaska has mandated labels for GM fish if they go on sale.



JMA/EUMETSAT

## Record storm ravages Philippines

One of the most powerful tropical cyclones on record, Super Typhoon Haiyan, devastated the central Philippines when it made landfall on 8 November. By some estimates, wind speeds topped 310 kilometres per hour as the storm reached the coast, placing Haiyan at or near the top of the list of strongest known storms. The storm surge destroyed much of the coastal

city of Tacloban. Buildings collapsed, and flooding ravaged a region still recovering from a magnitude-7.1 earthquake last month. Officials estimate that at least 10,000 people have died, giving Haiyan a second grim title: the deadliest storm in Philippine history. In 2012, Super Typhoon Bopha killed nearly 2,000 people in the southern part of the country.

### Ivory crush

The US Fish and Wildlife Service will destroy its stockpile of contraband elephant ivory on 14 November, officials announced last week. Over the past 25 years, US law enforcement has seized about 6 tonnes of illegal ivory. The tusks will be crushed to send a message to ivory traffickers worldwide, the agency says. In Tanzania, police recently seized nearly 800 elephant tusks in anti-poaching raids, according to media reports last week.

### Animal research

A set of draft guidelines committing to less secrecy on animal experiments, and signed by more than 50 organizations involved in UK life-sciences research, was

released for public comment last week (see [go.nature.com/3rqcn4](http://go.nature.com/3rqcn4)). Major research funders and institutions, coordinated by the London-based group Understanding Animal Research, say that they want to be more open about how animals are used in research. But anti-vivisection campaigners said that those backing the agreement did not want true openness about "painful and unscientific" animal experiments.

### Deep-sea trawling

European politicians voted to ban deep-sea bottom trawling in the Atlantic Ocean last week, but only in areas where ecosystems are deemed especially vulnerable. The move disappointed campaigners who have

pushed for a total ban on this method of fishing, which they say is highly damaging to certain species. The package of regulations approved by the European Parliament's fisheries committee will be considered by the full parliament in December.

### GM crop stir

A third genetically modified (GM) crop may soon be approved for cultivation in the European Union. The European Commission on 6 November asked member-state ministers to decide whether to authorize GM maize (corn) made by DuPont Pioneer in Johnston, Iowa, which first requested approval in 2001. A minority of ministers has consistently blocked approvals in the past,

PLC/BP so the commission had given up referring cases, but its hand was forced by a court ruling in September after DuPont sued. Only a majority of votes in favour of rejection will prevent the maize from being approved.

## PEOPLE

## Poisoning puzzles

Scientists at the University Centre of Legal Medicine in Lausanne, Switzerland, announced on 6 November that they had found traces of the radioactive metal polonium-210 in the exhumed body of Yasser Arafat, former president of the Palestinian National Authority. But because Arafat died nine years ago it is hard to distinguish levels of synthetic polonium from background radiation, so the results do not clearly establish that Arafat was poisoned, nuclear physicists told *Nature*. Separately, a forensic analysis of exhumed Chilean poet Pablo Neruda, released on 8 November, found no evidence that he was poisoned. See [go.nature.com/mpdrjd](http://go.nature.com/mpdrjd) and [go.nature.com/5jgggr](http://go.nature.com/5jgggr) for more.

## Energy nomination

US President Barack Obama has nominated physicist Ellen Williams (pictured) to head the Advanced Research Projects Agency-Energy, a branch of the Department



of Energy that funds high-risk, high-pay-off research. Currently on leave from the University of Maryland in College Park, Williams has been chief scientist for oil-and-gas giant BP since 2010. From 1996 to 2009, she directed the University of Maryland Materials Research Science and Engineering Center.

## Climate politics

Republican Kenneth Cuccinelli lost his bid on 5 November to become the next governor of Virginia, in an election that highlighted scientific issues. As the state's attorney-general, in 2010 Cuccinelli launched an investigation of climate scientist Michael Mann. Virginia's Supreme Court later ruled that he lacked the authority to demand Mann's grant applications, e-mails and other documents. During the gubernatorial campaign, Cuccinelli's opponent, Democrat Terry McAuliffe, called him a climate-change

denier and accused him of running a 'witch-hunt' against Mann.

## RESEARCH

## Genome access

An effort to encourage people to make their genome sequences and medical histories public has expanded from the United States to the United Kingdom. George Church, a genomicist at Harvard Medical School in Boston, Massachusetts, who launched the Personal Genome Project in 2007, and Stephan Beck, a genomicist at University College London, announced the UK arm on 6 November; a European franchise is on the way in 2014. See [go.nature.com/izmgpo](http://go.nature.com/izmgpo) for more.

## Greenhouse gases

Atmospheric greenhouse-gas concentrations reached a record high in 2012. According to a 6 November report by the World Meteorological Organization (WMO) in Geneva, Switzerland, carbon dioxide levels climbed last year to an average 393.1 parts per million (p.p.m.) — 141% above pre-industrial levels. Long-lived greenhouse gases — including CO<sub>2</sub>, methane, nitrous oxide and chlorofluorocarbons — rose collectively by 2.6 p.p.m. to an equivalent CO<sub>2</sub>

## COMING UP

## 18 NOVEMBER

NASA's MAVEN mission to Mars is scheduled to launch. See page 178 for more.

## 19–21 NOVEMBER

In Paris, the International Energy Agency holds its biennial meeting to set new strategic priorities.

[go.nature.com/qrijmcg](http://go.nature.com/qrijmcg)

concentration of 475.6 p.p.m.. The observations come from the WMO's global monitoring network, which last May measured a record daily concentration of atmospheric CO<sub>2</sub> at Mauna Loa, Hawaii (see *Nature* 497, 13–14; 2013).

## Biology preprints

Life scientists have a new website for sharing research papers before journal publication. The site, called bioRxiv, was started by Cold Spring Harbor Laboratory Press in New York, and went live on 11 November. Unlike the arXiv.org preprint server popular among physicists, the site will allow readers to comment on articles, and will include sections devoted to different subdisciplines of biology. See page 180 for more.

## BUSINESS

## Rare-disease deal

Pharmaceutical giant Shire, headquartered in Dublin, announced on 11 November the purchase of drug company ViroPharma for US\$4.2 billion. Based in Exton, Pennsylvania, ViroPharma develops drugs for rare diseases. The bulk of its roughly \$400 million in annual sales revenue comes from a drug used to treat a rare blood condition known as hereditary angioedema.

➔ [NATURE.COM](http://NATURE.COM)

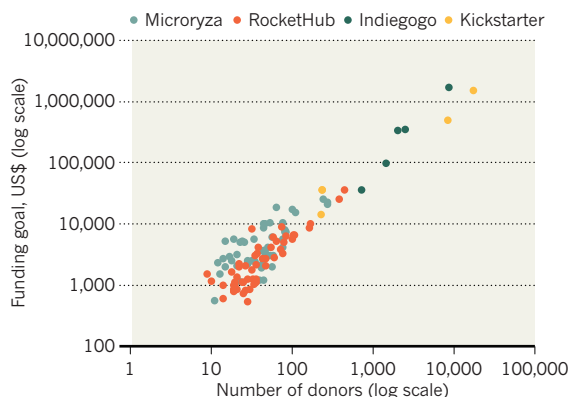
For daily news updates see: [www.nature.com/news](http://www.nature.com/news)

## TREND WATCH

Science projects seeking crowdfunding in the past two years have included a glow-in-the-dark plant that raised US\$484,013 (see *Nature* 498, 15–16; 2013) and an asteroid-hunting telescope called Arkyd that raised \$1.5 million. But an analysis of 115 science appeals shows that the median amount raised is about \$3,000, with mean donations of \$60–100. The data were collected by Ethan Perlstein, an evolutionary pharmacologist in Oakland, California (see [go.nature.com/3fembu](http://go.nature.com/3fembu)).

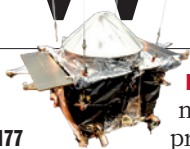
## THE SCIENCE OF CROWDFUNDING

On four popular crowdfunding websites, 115 successful science appeals collected more than US\$5 million from 2011 to 2013.



# NEWS IN FOCUS

**PHYSICS** Post-Higgs dreams of a 'very large' collider to dwarf LHC **p.177**

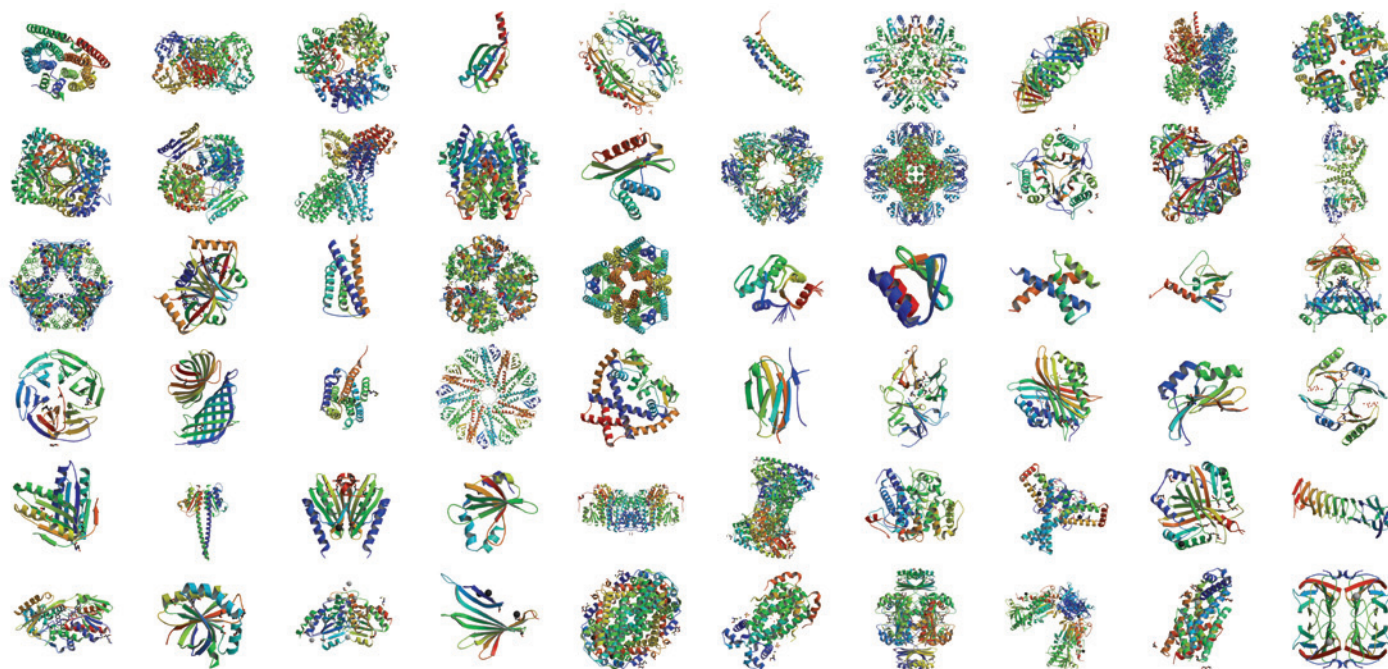


**PLANETARY SCIENCE** Mars mission MAVEN prepares for launch **p.178**

**US POLITICS** Social scientists rally in defence of their work **p.179**

**AIDS IN AFRICA** An ambitious push to circumcise 20 million men **p.182**

MARCA A. ELSLIGER/WWW.JCSG.ORG



A sample of the thousands of protein structures solved with the help of a US National Institutes of Health programme that has now been closed down.

## POLICY

# Large NIH projects cut

*Budget woes force institutes to scrutinize expensive, non-competitive programmes.*

BY SARA REARDON

Big science is under big pressure at the US National Institutes of Health (NIH). Gone are the glory days of the early 2000s, when a doubling of the agency's budget over five years allowed it to establish dozens of programmes with their own large, dedicated budgets. Since March, the mandatory US government cuts known as sequestration have sliced 5% off the NIH's already-tight budget. Now, institute directors are assessing the future of these costly legacy programmes — and shutting some of them down. Given the climate, directors are wondering whether their limited money might be better spent funding competitive grant applications, more than 80% of which are currently denied.

"In challenging fiscal times, it's incumbent on us to take a step back and look at how we're

investing money," says Jon Lorsch, who took over the National Institute of General Medical Sciences (NIGMS) in Bethesda, Maryland, in August. He says that the budget crunch has contributed to his decisions on what to cut, and that his institute's priority should be to fund research proposed by investigators, rather than programmes designed by the NIH.

Changes are already under way. In September, the NIGMS announced that it will wind down the Protein Structure Initiative (PSI), which since 2000 has received about US\$70 million per year in dedicated funds to develop ways to crystallize proteins and predict their structures. And on 1 November, the institute announced that its Pharmacogenomics Research Network, which studies how genes influence the effectiveness of drugs, will lose the protected budget that has brought it \$404 million over its 13-year lifetime.

An advisory council to the NIGMS had concluded that the PSI was successful but had run its course. And the field of pharmacogenomics, which was new when the research network began in 2000, has matured enough for its scientists to compete effectively with grant applicants in other fields, says Lorsch. "Now it's time to let those things fly and be free."

For investigators who feel that the NIH should hand out more small grants, the NIGMS's example is welcome. The PSI, in particular, has been controversial. In its first phase, the money was distributed between nine structural-biology research consortia, each of which would go on to produce dozens of protein structures per year — with the lofty goal of eventually predicting structures for most of the amino-acid sequences possible in nature.

Steven Almo, a crystallographer at the Albert Einstein College of Medicine in New





► York who leads one of these consortia, says that the programme sped up the development of technologies such as robotic crystallization systems and automated high-throughput structure screening. He adds that all structural biology benefits from these advances, as well as from the thousands of solved protein structures that have been deposited in the PSI's Structural Biology Knowledgebase (Nature Publishing Group partners with the NIGMS to maintain this database).

But many of the structures produced by the PSI, which were based on theoretical models, turned out to be irrelevant to biological functions, counters Gregory Petsko, a crystallographer at Brandeis University in Waltham, Massachusetts. If the PSI's budget had instead funded a few hundred individual investigators, each competing to work on specific structures, they probably would have come up with more relevant proteins, he says. He is thrilled that the programme is shutting down. "Put a stake through its heart, bury it in a coffin filled with its native soil — do whatever you can do to keep it from rising again," he says.

Partly in response to such criticisms, the PSI changed its direction in 2010, starting a new phase: PSI:Biology, which directed four centres to work on structures that biologists believe to be particularly important. Almo's group, for instance, is now investigating proteins involved in the immune system, to look for molecules that could be useful as drugs or drug targets. He is disappointed that the PSI is shutting down just

as the new phase is becoming productive. "We have a unique resource and we're in the position to start applying this in a powerful way," he says.

But Stefano Bertuzzi, executive director of the American Society for Cell Biology in Bethesda, says that the NIGMS has made the right decision. He thinks that it is better to make strategic cuts that eliminate entire programmes than to trim all projects and debilitate everyone.

That approach also suits Story Landis, director of the National Institute of Neurological Disorders and Stroke (NINDS) in Bethesda, which has already instigated such changes. In 2009, NINDS

could guarantee funding for only the top 9% of grant applications it received. So, Landis says, the institute began taking a close look at big, long-running legacy projects that had begun around 2000, and ended about 15 that were unproductive or outdated. This has saved the institute some \$30 million in the past four years. Ending ineffective clinical trials saved another \$92 million. NINDS now guarantees funding for 14% of grant applications.

"The issue wasn't big science, it's more making sure that we manage our dollars effectively," says Landis. Indeed, NINDS will not be abandoning big projects any time soon: the controversial BRAIN Initiative, which aims to develop technologies to map neural circuits,

could draw \$40 million in NIH money in 2014 — much of it from NINDS.

Still, a tight budget does seem to be enforcing higher standards among big projects being started across the NIH (see D. Sarewitz *Nature* **502**, 595; 2013). That is as it should be, says former NIGMS director Jeremy Berg, now director of the Institute of Personalized Medicine at the University of Pittsburgh in Pennsylvania. Berg does not believe that more money necessarily produces better science: in 2010, he performed an analysis which found that middle-sized labs with moderate funding are the most productive (see *Nature* **468**, 356–357; 2010).

Yet Berg does not want big science to disappear. There will always be problems that can be solved only by large initiatives, he says — such projects just have to have clear and relevant goals. "My strong feeling is that the NIH's goal is to try to facilitate the NIH mission, which is improving the nation's health," he says. "It's not a jobs programme for scientists."

Lorsch says that more changes can be expected at the NIGMS in the coming year, but he stresses that it will not be a battle between big science and little labs. "People set up a dichotomy of investigator-initiated projects versus team science, but that's not the contrast we're making," he says. And there could be room for growth in a medium-sized category. Lorsch says that the NIGMS is looking at new funding mechanisms, such as grants with multiple principal investigators and 'supergroups' composed of scientists from different fields. ■

## CLIMATE CHANGE

# Warsaw talks to thrash out UN climate roadmap

*Costs of reducing emissions may be flashpoints in path towards 2015 Paris treaty.*

BY JEFF TOLLEFSON

At a major United Nations climate summit in Warsaw this week, a plan is being hammered out for negotiations on a new climate treaty to be finalized in Paris in two years' time. Delegates from 195 nations are also seeking to obtain commitments from countries to limit their greenhouse-gas emissions between now and 2020. But the path forward is rife with disputes between rich and poor countries over funding, and how to allocate and enforce emissions reductions.

The conference aims to outline the schedule and to set parameters for negotiations ahead of the next major climate summit in Paris in 2015, when countries hope to forge a treaty to follow

the 2009 agreement settled on in Copenhagen.

At that meeting, negotiations over a formal treaty broke down, but eventually resulted in a set of non-binding pledges — the Copenhagen Accord — for emissions reductions until 2020. The accord also blurred the distinction between developed countries, which were bound by the 1997 Kyoto Protocol to reduce emissions, and developing countries, which had no such obligations. Since then, negotiators have worked on how to structure a new framework that would involve climate commitments from all countries — including China, now the world's largest emitter, and the United States, which never ratified the Kyoto Protocol (E. Diringer *Nature* **501**, 307–309; 2013).

The Warsaw talks are split into two main

tracks. One focuses on the architecture of a new global climate treaty that would take effect after 2020, when the current Copenhagen commitments expire. The second examines what can be done to strengthen commitments between now and 2020 to increase the chance of limiting global warming to a target of 2 °C above pre-industrial temperatures (see 'Emissions up in the air?').

The European Union (EU), for example, has proposed a multi-stage process, whereby commitments for climate action post-2020 would be registered next year and then subjected to an international assessment to determine how well the commitments measure up against each other and against scientific assessments. The final commitments would

► York who leads one of these consortia, says that the programme sped up the development of technologies such as robotic crystallization systems and automated high-throughput structure screening. He adds that all structural biology benefits from these advances, as well as from the thousands of solved protein structures that have been deposited in the PSI's Structural Biology Knowledgebase (Nature Publishing Group partners with the NIGMS to maintain this database).

But many of the structures produced by the PSI, which were based on theoretical models, turned out to be irrelevant to biological functions, counters Gregory Petsko, a crystallographer at Brandeis University in Waltham, Massachusetts. If the PSI's budget had instead funded a few hundred individual investigators, each competing to work on specific structures, they probably would have come up with more relevant proteins, he says. He is thrilled that the programme is shutting down. "Put a stake through its heart, bury it in a coffin filled with its native soil — do whatever you can do to keep it from rising again," he says.

Partly in response to such criticisms, the PSI changed its direction in 2010, starting a new phase: PSI:Biology, which directed four centres to work on structures that biologists believe to be particularly important. Almo's group, for instance, is now investigating proteins involved in the immune system, to look for molecules that could be useful as drugs or drug targets. He is disappointed that the PSI is shutting down just

as the new phase is becoming productive. "We have a unique resource and we're in the position to start applying this in a powerful way," he says.

But Stefano Bertuzzi, executive director of the American Society for Cell Biology in Bethesda, says that the NIGMS has made the right decision. He thinks that it is better to make strategic cuts that eliminate entire programmes than to trim all projects and debilitate everyone.

That approach also suits Story Landis, director of the National Institute of Neurological Disorders and Stroke (NINDS) in Bethesda, which has already instigated such changes. In 2009, NINDS

could guarantee funding for only the top 9% of grant applications it received. So, Landis says, the institute began taking a close look at big, long-running legacy projects that had begun around 2000, and ended about 15 that were unproductive or outdated. This has saved the institute some \$30 million in the past four years. Ending ineffective clinical trials saved another \$92 million. NINDS now guarantees funding for 14% of grant applications.

"The issue wasn't big science, it's more making sure that we manage our dollars effectively," says Landis. Indeed, NINDS will not be abandoning big projects any time soon: the controversial BRAIN Initiative, which aims to develop technologies to map neural circuits,

could draw \$40 million in NIH money in 2014 — much of it from NINDS.

Still, a tight budget does seem to be enforcing higher standards among big projects being started across the NIH (see D. Sarewitz *Nature* **502**, 595; 2013). That is as it should be, says former NIGMS director Jeremy Berg, now director of the Institute of Personalized Medicine at the University of Pittsburgh in Pennsylvania. Berg does not believe that more money necessarily produces better science: in 2010, he performed an analysis which found that middle-sized labs with moderate funding are the most productive (see *Nature* **468**, 356–357; 2010).

Yet Berg does not want big science to disappear. There will always be problems that can be solved only by large initiatives, he says — such projects just have to have clear and relevant goals. "My strong feeling is that the NIH's goal is to try to facilitate the NIH mission, which is improving the nation's health," he says. "It's not a jobs programme for scientists."

Lorsch says that more changes can be expected at the NIGMS in the coming year, but he stresses that it will not be a battle between big science and little labs. "People set up a dichotomy of investigator-initiated projects versus team science, but that's not the contrast we're making," he says. And there could be room for growth in a medium-sized category. Lorsch says that the NIGMS is looking at new funding mechanisms, such as grants with multiple principal investigators and 'supergroups' composed of scientists from different fields. ■

## CLIMATE CHANGE

# Warsaw talks to thrash out UN climate roadmap

*Costs of reducing emissions may be flashpoints in path towards 2015 Paris treaty.*

BY JEFF TOLLEFSON

At a major United Nations climate summit in Warsaw this week, a plan is being hammered out for negotiations on a new climate treaty to be finalized in Paris in two years' time. Delegates from 195 nations are also seeking to obtain commitments from countries to limit their greenhouse-gas emissions between now and 2020. But the path forward is rife with disputes between rich and poor countries over funding, and how to allocate and enforce emissions reductions.

The conference aims to outline the schedule and to set parameters for negotiations ahead of the next major climate summit in Paris in 2015, when countries hope to forge a treaty to follow

the 2009 agreement settled on in Copenhagen.

At that meeting, negotiations over a formal treaty broke down, but eventually resulted in a set of non-binding pledges — the Copenhagen Accord — for emissions reductions until 2020. The accord also blurred the distinction between developed countries, which were bound by the 1997 Kyoto Protocol to reduce emissions, and developing countries, which had no such obligations. Since then, negotiators have worked on how to structure a new framework that would involve climate commitments from all countries — including China, now the world's largest emitter, and the United States, which never ratified the Kyoto Protocol (E. Diringer *Nature* **501**, 307–309; 2013).

The Warsaw talks are split into two main

tracks. One focuses on the architecture of a new global climate treaty that would take effect after 2020, when the current Copenhagen commitments expire. The second examines what can be done to strengthen commitments between now and 2020 to increase the chance of limiting global warming to a target of 2 °C above pre-industrial temperatures (see 'Emissions up in the air?').

The European Union (EU), for example, has proposed a multi-stage process, whereby commitments for climate action post-2020 would be registered next year and then subjected to an international assessment to determine how well the commitments measure up against each other and against scientific assessments. The final commitments would

then be registered in Paris in 2015. By getting countries to volunteer their climate commitments and comparing them in this way, the hope is that nations with unambitious targets might be shamed into strengthening them. The EU has also called for a review of pre-2020 commitments.

Tasneem Essop, who is tracking negotiations for the environmental group WWF in Cape Town, South Africa, says that these short-term commitments are crucial for pointing the world in the right direction. "The biggest challenge will be to ensure that emissions do peak within this decade," she says.

The cost of reducing emissions could be the first flashpoint in Warsaw. In Copenhagen, developed countries agreed to provide US\$30 billion in climate aid from 2010 to 2012, and to increase climate support to developing countries to \$100 billion annually by 2020. Although the short-term commitments were largely met, there is no clear plan for attaining the goal of \$100 billion a year. From emerging giants such as Brazil and China to poor countries in Africa, developing nations are demanding that wealthy countries ramp up funding and create a viable path to this goal.

With public coffers strapped, many developed nations are looking for other funding sources. One possibility is to place some type of levy on international aviation, which is being considered by the International Civil Aviation Organization in Quebec, Canada. The body has committed to craft an agreement by 2016 that could take effect by 2020.

Negotiators in Warsaw will haggle over how to finance and ultimately deploy climate aid through organizations such as the newly launched Green Climate Fund, based in Incheon, South Korea. Another flashpoint is the developing countries' demand for a 'loss and damage' mechanism to compensate poor countries irreparably harmed by climate change.

But the biggest questions will centre on the framework for the treaty in 2015. Before Copenhagen, the emphasis was on a treaty similar to the Kyoto Protocol that would lock in legally binding emissions reductions. In Copenhagen, the United States and other developed countries pushed for an alternative that would allow individual countries to register commitments, which would then be reviewed at an international level. Delia Villagrasa, a senior adviser for the European Climate Foundation in Brussels, says that the talks are moving towards this bottom-up approach, which would be combined with a formal review to assess commitments and identify ways to scale them up. The world could get its first hint of what such a system might look like as the talks wrap up next week.

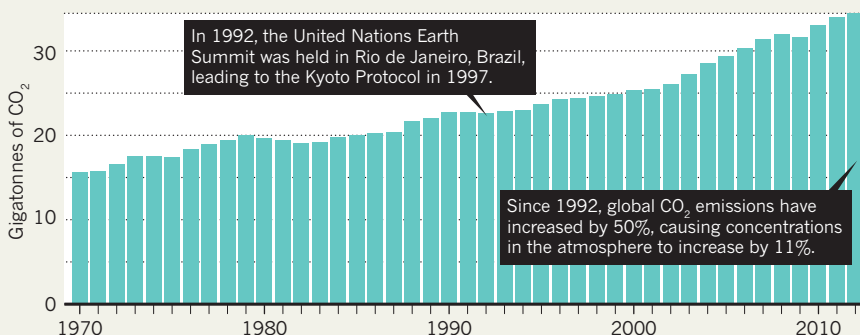
"Warsaw will bring some clarification on the structure of the new agreement," Villagrasa says. "That's not sexy for the media, but it's important." ■

Additional reporting by Quirin Schiermeier

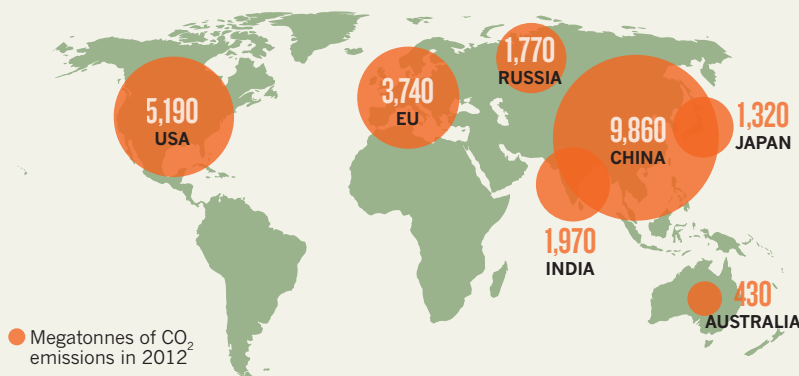
## EMISSIONS UP IN THE AIR?

Countries will discuss how to reduce greenhouse-gas emissions at this week's climate summit in Warsaw. Current commitments expire in 2020 and fall short of climate goals.

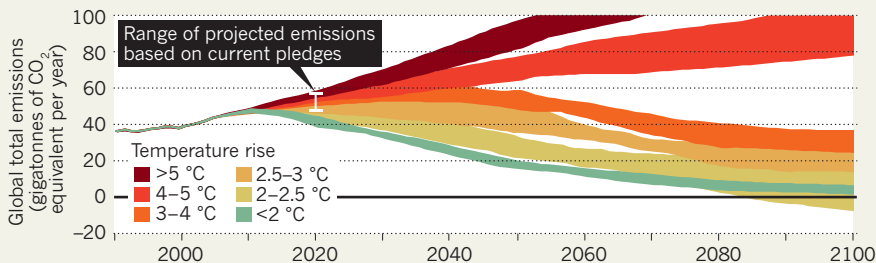
After rising by about 3% annually in the past decade, global emissions of carbon dioxide increased by just 1.1% in 2012.



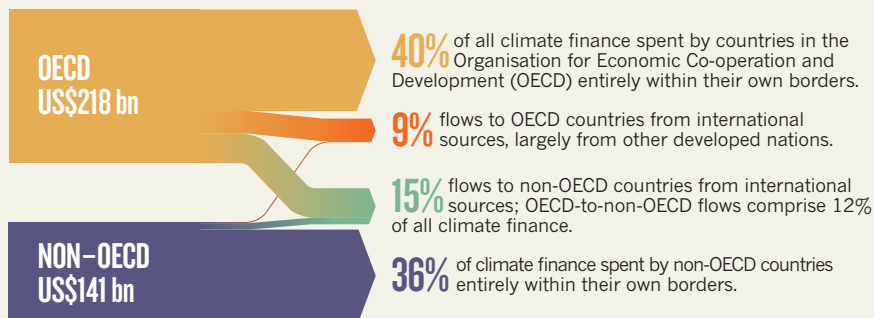
Developing countries, led by China, now emit more greenhouse gases than developed countries, although their per-capita emissions are much lower.



Various trajectories for global emissions are depicted according to their resulting temperature rises by 2100. Current commitments to reduce emissions by 2020 are not sufficient to limit warming to 2 °C.



Global climate investments by governments, businesses and consumers topped an estimated US\$359 billion in 2012. Some \$39 billion to \$62 billion, including \$4 billion to \$11 billion in government aid, moved from wealthy to poor countries.







Discoveries at the Large Hadron Collider, now being upgraded, will inform designs for the next accelerator.

# PARTICLE PHYSICS

# Physicists plan to build a bigger LHC

*Accelerator ring would be 100 kilometres around and run at seven times the energy of the Large Hadron Collider.*

BY EUGENIE SAMUEL REICH

When Europe's Large Hadron Collider (LHC) started in 2008, particle physicists would not have dreamt of asking for something bigger until they got their US\$5-billion machine to work. But with the 2012 discovery of the Higgs boson, the LHC has fulfilled its original promise — and physicists are beginning to get excited about designing a machine that might one day succeed it: the Very Large Hadron Collider (VLHC).

"It's only prudent to try to sketch a vision decades into the future," says Michael Peskin, a theoretical physicist at SLAC National Accelerator Laboratory in Menlo Park, California, who presented the VLHC concept to a US government advisory panel on 2 November.

The giant machine would dwarf all of its predecessors (see 'Lord of the rings'). It would collide protons at energies around 100 teraelectronvolts (TeV), compared with the planned 14 TeV of the LHC at CERN, Europe's particle-physics lab near Geneva in Switzerland. And it would require a tunnel 80–100 kilometres around, compared with the LHC's 27-km circumference. For the past decade or so, there has been little research money available worldwide to develop the concept. But this summer, at the Snowmass meeting in

Minneapolis, Minnesota — where hundreds of particle physicists assembled to dream up machines for their field's long-term future — the VLHC concept stood out as a favourite.

Some physicists caution that the VLHC would be only a small part of the global particle-physics agenda. Other priorities include: upgrading the LHC, which shut down in February for two years to boost its energies from 7 TeV to 14 TeV; plans to build an International Linear Collider in Japan, to collide beams of electrons and positrons as a complement to the LHC's proton findings; and a major US project to exploit high-intensity neutrino beams generated at the Fermi National Accelerator

Laboratory in Batavia, Illinois. Jonathan Rosner, a particle physicist at the University of Chicago, Illinois, who convened Snowmass, says that these forthcoming projects should be the focus. "It's premature to highlight the VLHC," he says.

In some ways, the interest in the VLHC is a sign that particle physicists are returning to their roots, pushing to ever higher energies to find the fundamental building blocks of nature.

They will have to justify it, however. The discovery of the Higgs particle lends support to the idea that some particles have mass because they interact with a pervasive, treacle-like Higgs field. Yet many aspects of the discovery are still not understood, including why the mass of the Higgs particle is so large. One way of explaining its heaviness is through supersymmetry theory, in which known particles are coupled with heavier ones that might be observed in bigger particle colliders. Although the LHC has not detected any signs of supersymmetry, Peskin hopes that a hint may come before the end of the decade, which would help to inform the design of a larger machine.

One advocate of a bigger machine is Nima Arkani-Hamed, a theoretical physicist at the Institute for Advanced Study in Princeton, New Jersey. In December, he will help to launch an institute in Beijing called the Center for Future High Energy Physics. Part of its explicit mission, he says, is to explore the physics that a future proton collider might investigate. William Barletta, an accelerator physicist at the Massachusetts Institute of Technology in Cambridge, says that this work is crucial to identify a machine size that will maximize the science per dollar. "We won't just give hand-waving arguments," he says.

To build a 100-TeV machine, Barletta adds, physicists will need to develop superconducting magnets that can operate at higher fields than the current generation, perhaps 20 tesla instead of 14 tesla. One leading candidate material for such magnets is niobium tin, which can withstand higher fields but is expensive and must be cooled below 18 kelvin.

CERN is developing its own plans for a collider that is similar to the VLHC. CERN accelerator physicist Michael Benedikt is leading a study of a 'very high energy large hadron collider' that would pass under Lake Geneva. It would have the same key parameters as the suggested VLHC: a circumference of 80–100 km and a collision energy of 100 TeV. Benedikt suggests that construction might begin in the 2020s so that the machine could be completed soon after the LHC shuts down for good around 2035. "One would not want to end up with a huge gap for high-energy physics," he says. He adds that it is too early to offer a price tag. But other physicists speculate that a next-generation collider would have to cost less than \$10 billion for the project to be politically plausible. ■

## LORD OF THE RINGS

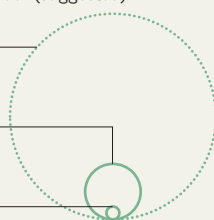
Physicists are discussing a proton-colliding machine that would dwarf the energy of its predecessors.

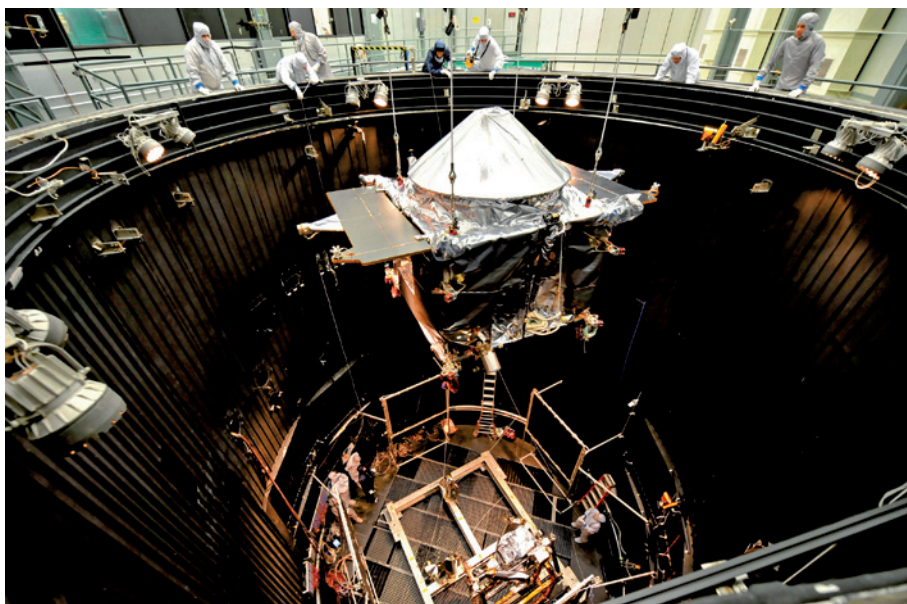
Very Large Hadron Collider (suggested)  
100 km  
100 TeV\*

Large Hadron Collider  
27 km  
14 TeV

Tevatron (closed)  
Circumference: 6.3 km  
Energy: 2 TeV

\*TeV, teraelectronvolt.





NASA's MAVEN spacecraft will measure the effect of solar storms on Mars' thin outer atmosphere.

PLANETARY SCIENCE

# Mars mission set for launch

*NASA's MAVEN orbiter aims to unravel the mystery of the red planet's missing atmosphere.*

BY ALEXANDRA WITZ

Just before he submitted a proposal for a mission to Mars to NASA, Bruce Jakosky checked to make sure the word 'aeronomy' hadn't crept into the document. Yes, his satellite would be studying the planet's wispy upper atmosphere, but he knew that NASA would be more interested in the links between the vanishing air and the planet's lost surface water. "We pitched it not as an aeronomy mission, but as a mission to understand the geology," says Jakosky, a planetary geologist at the University of Colorado Boulder.

The strategy worked. In 2008, NASA gave the Mars Atmosphere and Volatile Evolution (MAVEN) mission the go-ahead. The US\$485-million spacecraft is now at Florida's Cape Canaveral, waiting for a 28-day launch window that opens on 18 November. If MAVEN successfully reaches Mars ten months later, it will study how gases in the outer atmosphere escape into space — and how that rate of loss has changed over the past few billion years.

MAVEN comes at a crucial time in NASA's Mars programme. The spacecraft is the only US orbiter planned for the foreseeable future,

and it carries the long-term burden of relaying data from NASA rovers on the surface to mission controllers on Earth. The two operating US orbiters, Mars Odyssey and the Mars Reconnaissance Orbiter, were launched in 2001 and 2005, respectively. Of the three, MAVEN is the most likely to be working by the time NASA's next rover heads to Mars in 2020. The MAVEN project was spared from last month's US government shutdown because of this crucial relay ability.

Upon arrival, MAVEN will explore how the solar wind interacts with Mars's upper atmosphere, helping to erode it. "MAVEN's job is to obtain some key constraints that will allow us to make much more solid arguments about how Mars got to be what it is," says Janet Luhmann, a space physicist at the University of California, Berkeley, and the mission's deputy principal investigator.

Canyons carved into the landscape and chemical evidence from rocks indicate that Mars once had a thick atmosphere that kept the surface warm enough for water to flow. But most of that atmosphere is gone; today it is only about 1% as dense as that of Earth. Mars is so small that the gravitational grip it has on

its atmosphere is weak. And when its protective global magnetic field switched off, around 4 billion years ago, the planet's hold on its outer atmosphere weakened further.

MAVEN's instruments will explore several of the many ways that atoms in the atmosphere can escape. An ultraviolet spectrograph will measure major components of the atmosphere, including the ratio of deuterium (or 'heavy hydrogen') to hydrogen. The Martian atmosphere is enriched in deuterium because, unlike lighter-weight hydrogen, the isotope had insufficient thermal energy to break free of the planet's gravitational pull. Comparing the ratio of deuterium to hydrogen in the atmosphere with the ratio locked in ancient rocks at the planet's surface should reveal the rate of this particular type of loss over time.

Other missions to Mars have studied another type of loss: the depletion of charged particles. In 1988, the Soviet Phobos-2 mission made the first detailed measurements showing that the solar wind stripped ions from Mars's atmosphere. By 2007, the European Space Agency's Mars Express orbiter had discovered that several grams of oxygen and carbon dioxide ions were being lost each second (S. Barabash *et al. Science* **315**, 501–503; 2007).

Unlike these missions, MAVEN will be able to study a third type of escape: the erosion of neutral particles, such as hot oxygen atoms, that form in the upper atmosphere and escape because of their high energies. In addition, MAVEN will be able to study changes in extreme ultraviolet light flooding from the Sun, and watch for differences in the solar wind caused by solar storms, to measure the effect of solar activity on atmospheric losses. Together, these measurements should paint a more detailed picture of atmospheric escape than ever before, says Jakosky.

Mars Express will also continue to contribute. Its researchers and those working on MAVEN are planning to make simultaneous observations with the two spacecraft. When both are in place, their elliptical paths will take them from a few hundred kilometres above the planet's surface to thousands of kilometres away. While one of the probes is deep in the atmosphere, the other will have a broader perspective from afar. "Having two spacecraft capable of focusing on the same subject is very exciting," says Stas Barabash at the Swedish Institute of Space Physics in Kiruna, a member of the Mars Express science team.

Mars Express has funds to operate until the end of 2016. India's first Mars mission, a methane-sniffing orbiter that launched on 5 November, will also overlap with MAVEN: both spacecraft are due to arrive at Mars in the same month.

MAVEN has a one-year nominal mission in which to accomplish its goals. But as it fulfils its longer-term communications tasks through the years, it may be well placed to watch the Martian atmosphere eroding before its electronic eyes. ■

LOCKHEED MARTIN



## POLITICS

# Social scientists hit back at grant rules

*Researchers seek to fend off restrictions on National Science Foundation grant programmes.*

BY SARAH ZHANG

It is not unusual for conservative politicians in the United States to question the value of social-science research. Studies of anything from global social networks to the history of conservation in South America have proved irresistible to Republicans keen to argue that funding would reap greater rewards elsewhere. But this year, researchers in the field received a sharp shock when those criticisms morphed into tangible restrictions.

“What’s different this time is they succeeded,” says Howard Silver, executive director of the Consortium of Social Science Associations. In March, Congress placed new limits on political-science research funded by the National Science Foundation (NSF). And with powerful Republicans pushing to enact legislation that would restrict the NSF further, forcing it to support only research that serves the ‘national interest’, newly emboldened social scientists are gearing up for a fight.

Silver estimates that he has made about 25 visits to Capitol Hill since March to try to ensure that the NSF is still able to support a wide range of social, behavioural and economic (SBE) research. And SAGE Publications, a major publisher of social-science journals based in Thousand Oaks, California, has this year poured some US\$90,000 into lobbying for sustained NSF funding.

Researchers say that the March bill is already biting. The law includes a provision inserted by Senator Tom Coburn (Republican, Oklahoma) requiring the NSF to certify that all political-science projects it supports will benefit national security or US economic interests. In July, the NSF made a last-minute decision to scrap one of its two annual calls for political-science grants. The agency has not explained why, but many researchers blame Coburn’s amendment.

“This is catastrophic intervention in peer review,” says Jane Mansbridge, a political scientist at Harvard University’s John F. Kennedy School of Government in Cambridge, Massachusetts, who studies the democratic process. Mansbridge adds that she will not apply for funding in the NSF’s next grant call, in January, because she does not think that her proposal on the dynamics of negotiations will meet the Coburn amendment’s criteria.

Coburn’s success seems to have heartened other social-science critics in Congress. In April, Representative Lamar Smith (Republican, Texas), the chairman of the House Committee on Science, Space, and Technology, asked the NSF for details of five social-science grants, citing concerns that they did not satisfy the agency’s standard for ‘intellectual merit’. And Smith’s committee is this week holding a hearing to discuss legislation he is backing that would require the NSF to certify that all the grants it awards serve the national interest in one of six areas: economic competitiveness, health and welfare, scientific literacy, partnerships between academia and industry, promotion of scientific progress or national defence.

It is not just the idea of national interest that is concerning lobbyists. Smith’s bill does not include budget numbers, normally the centrepiece of authorization legislation. But it would

**“This is catastrophic intervention in peer review.”**

prevent NSF divisions other than the SBE directorate from funding social-science research unless it is of ‘higher priority’ than all other research. That has left many wondering whether Republicans ultimately plan to cut the SBE directorate’s budget.

Others say that the requirement to serve the national interest could steer the NSF away from basic research. “They’re sending a message to the NSF: ‘We want you to change your mission,’” says Michael Lubell, director of public affairs for the American Physical Society.

In their visits to Capitol Hill, lobbyists are stressing to lawmakers and their staff the far-reaching effects of NSF-funded research. They cite, for example, an algorithm that is now used to match donated kidneys to transplant patients.

But David Stonner, who directed the NSF’s congressional-affairs office from 1991 to 2007, worries that social scientists are taking the wrong approach. They need to highlight private industries, such as social-media companies, where the economic benefits are realized, he says. Others worry that such a tack will only reinforce prejudice against basic research. Myron Gutmann, former head of the NSF’s social-science programmes, says simply: “The NSF is not the ‘National Industry Support Foundation.’” ■



# Preprints come to life

A dedicated website for sharing biology papers before peer review leaves journals divided.

BY EWEN CALLAWAY

What are biologists so afraid of? Physicists, mathematicians and social scientists routinely post their research to preprint servers such as arXiv.org before publication, yet few life scientists follow suit.

A website that goes live this week is hoping to change that. The site, bioRxiv.org, launched by Cold Spring Harbor Laboratory Press in New York, bills itself as “the preprint server for biology”. It will operate similarly to arXiv, with scientists depositing papers as soon as they are ready to share them, weeks or months before formal publication.

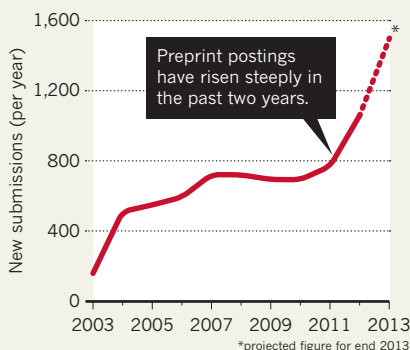
The site's launch comes as biologists begin to embrace preprints (see *Nature* 488, 19; 2012). ArXiv's quantitative biology section, launched in 2003, has seen exponential gains in the past two years (see ‘Life sciences freed’). It is on target to attract more than 1,500 manuscripts this year, says Paul Ginsparg, a physicist at Cornell University in Ithaca, New York, who started arXiv in 1991 and serves on bioRxiv's advisory board. Meanwhile, the open-access journal *PeerJ* launched its own biology preprint server in April and has drawn some 90 submissions so far.

Geneticists in particular have flocked to arXiv's quantitative-biology repository, and papers posted there on topics such as the genetic ancestry of southern Africans and Europeans have gone on to be published in prominent journals like *Nature Communications* and *PLoS Biology*. Scientists studying infectious diseases have also used the site to rapidly report on ongoing outbreaks, such as the H7N9 influenza infections in China in early 2013.

But arXiv does not plan to expand into other areas of the life sciences, such as cell and molecular biology, and the site is not designed with biologists in mind. Richard Sever, who founded bioRxiv with colleague John Inglis at Cold Spring Harbor Laboratory Press, says:

## LIFE SCIENCES FREED

The number of papers posted to the quantitative biology section of the preprint website arXiv.org has grown rapidly since 2003.



“This is not something that is competing with arXiv. It is supposed to complement it.”

BioRxiv, which operates in conjunction with HighWire Press in Stanford, California, is intentionally minimalist. “It's designed to be basic and functional because it's a mechanism for sharing manuscripts,” says Sever.

Papers — there are nine already, most of which had previously been posted to arXiv — will be grouped into some two dozen sub-disciplines, including cancer biology and immunology. Individual papers come with unique digital object identifiers (DOIs), making them easy to cite, and are described as new results, confirmatory results or contradictory results. Researchers can discuss papers in comments that appear alongside articles; these will be moderated.

At present, Cold Spring Harbor Laboratory is paying for the site, but Sever and Inglis declined to disclose the cost of the site's launch or its annual budget. They will look for ongoing sources of revenue, such as foundation grants, and plan to keep the site free to users. ArXiv's annual budget is around US\$750,000, a large proportion of which comes from the

Simons Foundation, a non-profit organization in New York, with universities and others making smaller contributions.

Leonid Kruglyak, a geneticist at the University of California, Los Angeles, and a member of bioRxiv's advisory board, expects many colleagues to adopt the site. Barriers that previously prevented biologists from using preprint servers seem to be eroding, he says. Most quantitative geneticists are no longer afraid of getting scooped by posting preprints of their work to arXiv, with many now seeing the process as a way to establish priority for their work.

Many journals, such as those published by the Genetics Society of America and the Ecological Society of America, have changed their policies in recent months to allow the practice. Nature Publishing Group journals allow preprints, and *Science* accepts them in many cases. Cell Press, which publishes biology stalwarts such as *Cell* and *Neuron*, discourages preprint posting. But “you can easily imagine a situation not too long from now, where having your journal not be preprint-friendly is an active disincentive to publishing there”, Kruglyak says.

However, Martin Fenner, a haematologist at Hanover Medical School in Germany, thinks that many life scientists will be more comfortable with open peer review, which is available from journals such as F1000Research. In open peer review, publishers post submitted draft manuscripts to a journal's website before acceptance. Compared with preprints posted to a third party such as bioRxiv, this process results in a shorter lag time between first draft and final publication, and the different versions are clearly linked, Fenner says.

Graham Coop, an evolutionary geneticist at the University of California, Davis, posts his lab's papers to arXiv and has already cross-posted one to bioRxiv. Peer pressure will prompt researchers in other fields to follow suit, he says. “If people see papers from people they know appearing on sites like the bioRxiv, they naturally feel more comfortable.” ■

SOURCE: ARXIV



**MORE  
ONLINE**

### TOP STORY



Hubble Space Telescope spots comet-like object with six tails  
[go.nature.com/avkvvs](http://go.nature.com/avkvvs)

### MORE NEWS

- Montreal Protocol helped with more than ozone hole [go.nature.com/g93gb6](http://go.nature.com/g93gb6)
- Method for predicting *h*-index may be flawed [go.nature.com/ymoeoe](http://go.nature.com/ymoeoe)
- Scant evidence Arafat and Neruda were poisoned [go.nature.com/mpdrjd](http://go.nature.com/mpdrjd) & [go.nature.com/5jqggr](http://go.nature.com/5jqggr)

### NATURE PODCAST



What to do when antibiotics fail; larger groups help social skills; and treating phonons like particles [nature.com/nature/podcast](http://nature.com/nature/podcast)

# Africa's circumcision CHALLENGE

*To combat the spread of HIV, health officials plan to circumcise 20 million men in Africa, but some have concerns about the aftermath.*

BY CATHERINE DE LANGE

**M**arvin is 22 years old, single and — like many men of his age — thinking about girlfriends. But for the next six weeks, he will have to give his love life a rest. On this sunny morning in the Zambian capital, Lusaka, Marvin is waiting to get circumcised. Along with three of his friends, he sits tensely on a bench outside the operating theatre, a room usually used for abortions. Are they nervous? Marvin cracks a wide smile. “A little,” he says.

In Lusaka, young men like Marvin have grown up with daily reminders of the risks that can accompany sexual relationships. For years, billboards and adverts painted on walls have espoused the importance of safe sex and testing for HIV, which currently infects more than one in five Lusaka residents. But more recently, a new message has been popping up in public spaces and the media: a call for men to get circumcised to reduce their chances of contracting the virus. Marvin's friends told him about the procedure. “A lot of them have been circumcised so I thought of doing the same thing,” he says, adding that it may increase his odds with women. “I hear that the first thing they ask you is ‘have you been circumcised?’” he says. But, ultimately, he hopes the surgery will help him “just to stay healthy”.

Scenes like this have become familiar across much of southern and eastern Africa. Since 2007, 14 nations have taken part in a massive public-health campaign aimed at circumcising millions of men in an attempt to drastically reduce the spread of HIV (see ‘Making the cut’). About 3 million men in the region have been circumcised since the start of the campaign, and the initiative was made a high priority in late 2011 by the World Health Organization (WHO), the Joint United Nations Programme on HIV/AIDS (UNAIDS) and the US President's Emergency Plan for AIDS Relief (PEPFAR) — which are funding part of the programme. The rest is being provided by the Bill & Melinda Gates Foundation, the World Bank and other global health organizations. The targets are ambitious: 80% of men of reproductive age in these countries need to be circumcised by 2015. That means more than 20 million circumcisions.

If these efforts succeed, the payoff could be considerable. Reaching the 80% goal could cut the number of new HIV infections in the target countries by as much as half, says Bertran Auvert, a public-health specialist at the University of Versailles Saint-Quentin-en-Yvelines in Paris. “The goal is to have a huge impact on the overall HIV epidemic in Africa.”

But some scientists worry that the benefits reported for circumcision

in clinical trials will not bear out to the same degree when scaled up to tackle a messy epidemic that is spread as much through behaviour as biology. In particular, they say that men are getting mixed messages about the benefits of circumcision. Another concern is the effect on women, who gain no direct protection from the circumcision campaign, and may even face greater risk. Critics say the programmes could increase the risk of HIV infection in some populations by encouraging people to engage in risky sexual behaviours, such as forgoing condoms.

“It's only been shown to decrease to some degree transmission from the female to the male, and yet we are acting as if it's the only thing to do,” says Philip Thuma, a doctor who runs the Malaria Institute at Macha in southern Zambia. “Our job as scientists is to step back and be a bit sceptical. Should we really be putting this much emphasis on something at the expense of other things that are just as important?”

## POINT OF ENTRY

The idea that circumcision might reduce the risk of HIV infection is often traced back to Valiere Alcena, a physician now at the Albert Einstein College of Medicine in New York City, who noted in 1986 that the men of Haiti and Central Africa are usually uncircumcised and as a result often develop lesions on the foreskin that serve as a potential entry point for AIDS<sup>1</sup>. The idea was cemented when a landmark study<sup>2</sup> in 1989 showed that of some 400 Kenyan men who visited a group of prostitutes with high rates of HIV, those who were not circumcised had more than eight times the risk of becoming infected than those who were. Michel Garenne, a demographer at the Pasteur Institute in Paris, remembers being at a meeting when the paper was first presented, and it seemed to make sense. “I was working in Senegal, where there was very little AIDS and a lot of circumcision, so I became quite convinced that something was going on,” he says.

But others disagreed, arguing that some populations did not fit this pattern and that the correlation could be explained by other factors relating to the prevalence of circumcision, such as religious practice. Garenne himself became sceptical. In 1994, he was working in South Africa, where he found a huge HIV epidemic even though many men were

**Circumcision packets at a clinic in Zambia.**

## ➔ NATURE.COM

For an online slideshow and audio content, see: [go.nature.com/dtgxyx](http://go.nature.com/dtgxyx)





circumcised. “I changed my mind and became convinced that circumcision will never help in controlling HIV/AIDS,” he says.

Those who did believe in the protective effect of circumcision called for a systematic analysis to settle the matter. Three randomized clinical trials were carried out in southern Africa between 2002 and 2007. One study<sup>3</sup>, led by Auvert, offered medical circumcision to men in a region of South Africa and compared them with a non-circumcised group. After 21 months, there were 20 cases of HIV in the circumcised group and 49 among the uncircumcised men, equating to a 60% reduction in HIV risk.

Two other trials, in Kenya<sup>4</sup> and Uganda<sup>5</sup>, produced similar results. “It’s extraordinary in any public-health measure that you get such consistency of results,” says Robert Bailey, an epidemiologist at the University of Illinois at Chicago, who led the Kenyan study.

The WHO also found the consistency compelling and declared the results “an important landmark in the history of HIV prevention”. In 2007, the WHO, UNAIDS and PEPFAR recommended implementing voluntary medical male circumcision (VMMC) in 14 high-risk African countries.

### ‘A MAN WHO CARES’

Zambia, with its high prevalence of HIV, low circumcision rates and mixed attitudes towards circumcision, is a microcosm of the broader campaign. During the past five years, it has seen the progressive roll-out of catchy, pro-circumcision slogans on posters, television and radio campaigns. Health-service providers have also developed outreach programmes to spread the message.

At the clinic where Marvin is waiting, visitors are greeted by an advertisement for its services: a silhouette of a man standing proudly with his hands hooked inside his belt and, above it, a slogan reading: “Male Circumcision, a man who cares”. Clinics that perform the procedure hope the belt-hook stance will become an unspoken symbol, a way for men to implicitly communicate their circumcised status, which is increasingly viewed as an attractive attribute.

If the campaigns can convince enough men to take part, the programme could significantly reduce HIV-transmission rates, according to a number of modelling studies. These project that circumcising 80% of men aged 15–49 years old in targeted countries by 2015 would decrease the incidence of HIV by 30–50% over 10 years, amounting to some 3.4 million fewer new infections. Carrying out these 20.3 million circumcisions by 2015, plus an additional 8.4 million over the following 10 years, would cost about US\$2 billion, but would potentially save \$16.5 billion by 2025 owing to avoided treatment and care costs<sup>6</sup>.

There may be other benefits too. Before men undergo voluntary medical circumcision, many get an HIV test, which means more cases of HIV will be picked up, and thus more men will be referred for treatment. “I think circumcision is going to have a very significant impact because it’s being delivered as a package,” Bailey says.

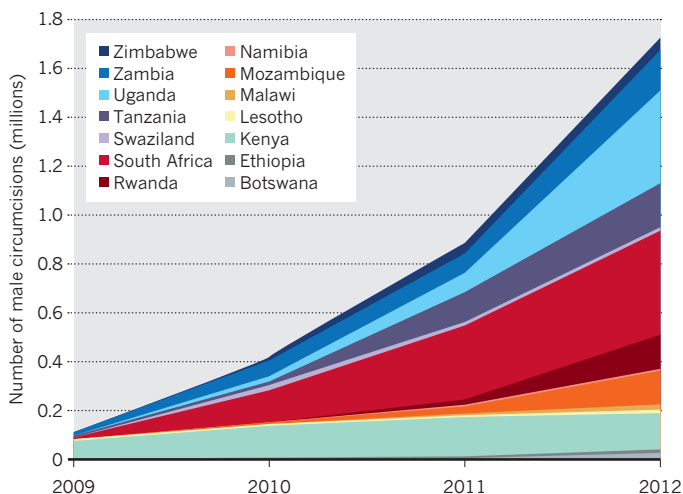
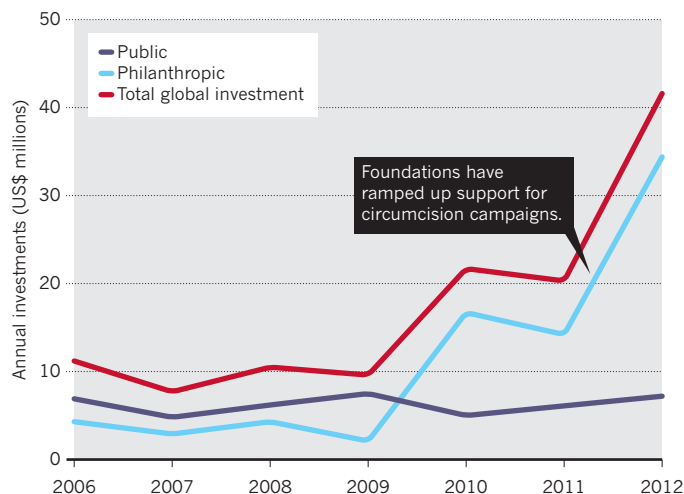
A big part of that package is a one-on-one counselling session run by people such as Jonathan Kabanda, a counsellor for the Society for Family Health in Livingstone, about five hours’ drive south of Lusaka. It is his job to make sure that men who come in for circumcision understand the benefits and the risks, not only for themselves but also for their partners. On a Saturday morning, Kabanda sits down and arranges his materials, among them three large wooden phalluses — one equipped with a retractable fabric foreskin — and some condoms. He pulls out a small flip chart that details different matters relating to circumcision — starting with how the foreskin can increase the transmission of HIV and leading on to the benefits and limitations of circumcision.

Those who support the campaigns, including the WHO, PEPFAR and the authors of the clinical trials, stress the importance of making sure that patients understand that circumcision provides limited protection and that men should continue to use condoms. But critics ask whether the message is getting through. There is certainly room for confusion. Although Kabanda’s presentation is thorough, one page of the flip chart, titled ‘MC does not offer 100% protection’ shows a picture of a man chasing two women down the street while throwing his condoms away.



## MAKING THE CUT

Since health trials in Africa demonstrated that circumcision lowers rates of HIV transmission, donor nations and charities have poured more than US\$100 million into circumcision campaigns in 14 African nations. But they have a long way to go to reach the goal of 20 million circumcisions by 2015.



Other benefits are overstated: “You are preventing your partner from cervical cancer,” Kabanda says, when, in fact, the reduction in risk is far from complete.

Outside the clinic in Lusaka, it is clear Marvin has also received mixed messages from his counselling session. “We were told there was a sixty-something per cent of being ... I don’t know if it’s from HIV or some sexually transmitted disease,” he says. “And the 45%, I don’t know what it is; I didn’t get much about that. I think I’ll try to ask later.”

Many men may face similar confusion. As part of the procedure, participants get only one counselling session on the matter of circumcision and one with their HIV test. By contrast, the men in the clinical trials had sessions before and after circumcision, and every six months during the 21-month follow-up period. Some public-health advocates argue that the protocols used in the trials should be replicated closely in the campaigns, otherwise they will not be as effective. “When technologies like male circumcision are demonstrated effective, everybody forgets about the rest of the protocol and just looks at the medical procedure,” says Seth Kalichman, a clinical psychologist at the University of Connecticut in Storrs who studies behavioural responses to HIV-prevention campaigns. “We need to think about the standard of care within which a procedure is implemented if we’re going to hope to get the effects that we saw in the trial.”

However, in a campaign that aims to reach millions of men, it would be hard to get participants to return for regular counselling. And some researchers, such as Auvert, dismiss the importance of counselling, arguing that it has little impact on behaviour. “If counselling was enough to change sexual behaviour to reduce the risk of acquiring HIV, we would have stopped the HIV epidemic in Africa a long time ago,” he says.

### RISK ASSESSMENT

The question of behaviour has become a crucial one in the debate over circumcision. A chief concern is that men and women who believe that they are less likely to acquire HIV will engage in more risky sexual behaviour. This means that circumcision “could have the opposite effect in the long run”, Garenne says. “People might be more likely to get HIV than if they were not circumcised at all.”

Studies of this phenomenon — known as risk compensation — are just starting to shed light on how circumcision alters behaviour, and so far the findings are mixed. In one survey<sup>7</sup> conducted in Kenya directly before the circumcision scale-up took place, Bailey and his colleagues reported that one-fifth of men and women they interviewed said that condom use is less necessary and that HIV is a less serious threat because male circumcision is available.

These misconceptions could put women at particular risk. Erika

Layer, a health-communications researchers at the Johns Hopkins University in Baltimore, Maryland, explored that possibility by interviewing 33 women following a circumcision campaign in the city of Iringa in Tanzania. She found an overwhelming preference for circumcised men<sup>8</sup>; women believed them to be cleaner and better educated. Some women also assumed — incorrectly — that the reduced risk of contracting HIV applied to women, too. As one woman who was interviewed for the study put it: “Circumcised men who are HIV-positive can have sex with women and the women will not get the infection easily.”

Much of the research on risk compensation relies on self-reported behaviours and often involves qualitative studies. But Layer’s interviews showed consistent misunderstandings about circumcision, a finding borne out by studies in other countries that suggest women are less likely to require circumcised men to use a condom. “There is evidence in South Africa, Kenya and Tanzania that shows that generally women are overstating the benefits of circumcision and believe the risk to be lower than it actually is,” she says.

Other studies, however, suggest that circumcision does not lead to more risky behaviour by men and women. Earlier this year, Auvert published findings from a survey<sup>9</sup> in the same region as his original trial in 2005. He found that despite a high uptake of circumcision, there was no reported difference in sexual behaviour — for instance in condom use or number of partners — between circumcised and uncircumcised men, and the estimated HIV incidence rate was around 60% lower among the circumcised group. Similarly, when Bailey and his colleagues did a follow-up study in Kenya, they found no increase in risk associated with the circumcision.

Researchers are at odds over the implications. “Any public-health measure will have the potential for risk compensation, but there is no evidence at all that risk compensation is taking place and indeed there is some evidence that the opposite is occurring,” Bailey says.

Kalichman disagrees. He believes that men are getting mixed messages, and that it doesn’t make sense that they would agree to circumcision if they were going to consistently use condoms, which offer a much higher level of protection. “Why bother with the procedure if you have to continue to use condoms?” he asks. “The whole point is, ‘if I have the procedure I don’t need to.’”

The take-home message on risk compensation, he says, is that “we don’t know the extent of it and we also don’t know the impact of it”. It is possible, he adds, that the overall effect is minor because the 60% risk-reduction offered by circumcision is already so substantial.

Nonetheless, Layer says it is important to make sure women are also targeted by clear messages about the limitations of circumcision. “If we do a little bit more to try to engage women I think we could reduce the

SOURCE: UNAIDS



Billboard adverts and family-planning programmes are helping to spread the message about the benefits of male circumcision.

PHILMON BULAWAYO/REUTERS; CATHERINE DE LANGE

immediate and the long-term risk for both women and men," she says.

Circumcision advocates say that in the long run, a general reduction in HIV prevalence in the population will ultimately benefit women, even if some women are placed at greater risk in the short term. "I am absolutely convinced that some women will become HIV positive because of the circumcision of their partners," says Auvert. However, he adds that he will publish some preliminary results soon showing that the impact on women is extremely small. "We are not concerned," he says.

The issue of balancing individual risks against widespread benefits also comes up in studies looking at how many men have sex during the healing period after circumcision, when open wounds and unhealed stitches increase the risk of transmission. In the counselling session, men are told to abstain from sex for six weeks or, if they can't wait, to use a condom. However, a study<sup>10</sup> by Paul Hewett of the Population Council in Lusaka and his colleagues found that around one-quarter of men said they resumed sex prematurely; of those, 82% engaged in at least one act of unprotected sex, and a number had unprotected intercourse with two or more partners. When the team modelled the effect of this on HIV transmission, it estimated that if 61,000 men were circumcised in one year, early resumption of sex would result in a total of 69 extra HIV infections. But the overall net effect was protective, with 230 fewer HIV infections in one year, predominantly among men.

### TARGET DRIVEN

Both critics and advocates of the circumcision campaigns agree that it will be a struggle to reach the 20-million mark by 2015, given that so far countries have only reached 15% of that overall goal. At health clinics across Zambia, much of the conversation is about targets, says Layer. "This seems to be the trend — do whatever it takes to get men in and out that door," she says. So circumcision providers are coming up with innovative ways to recruit more men. The US embassy in Zambia has launched a new effort to work with chiefs of traditionally non-circumcising tribes in the hope that they will "convince their own people that even though this isn't a part of their historic tradition this is the right choice," says Mark Storella, the US ambassador to Zambia. And mobile clinics travel to hard-to-reach areas to recruit men. "There's a lot of pressure right now in this country because of PEPFAR and others," says Thuma.

The recruitment drive is happening across the region. In Kenya, some service providers are even offering 'moonlight circumcisions' at night,

or financial incentives for referrals. More money is also being invested, but this has put significant pressure on local service providers to operate on more men. The Bill & Melinda Gates Foundation, based in Seattle, Washington, has implemented a pay-for-performance model in Zambia, for instance, in which organizations are reimbursed on the basis of their ability to reach a target number of circumcisions. "I think you have to be concerned about the impact on quality of such an incentives approach," says one person familiar with the programme who wished to remain anonymous. He wonders whether it could cause circumcision providers to cut costs for other essential services, such as counselling. In a statement to *Nature*, Sema Sgaier, a senior programme officer with the Gates Foundation, wrote "to ensure that VMMC delivery does not affect estab-

lished standards of care, it is always delivered as one component of a comprehensive package of services". These include HIV counselling and testing as well as screening and treatment for other sexually transmitted diseases.

Everybody from donors to doctors in the clinic agrees that the circumcision effort must be part of a broader strategy to strangle Africa's HIV epidemic. And although the 20-million goal remains distant, the campaigns could still bring major benefits, says Bailey. "Even if you don't reach the 80% target that's not to say you're

not going to have a large impact on the number of new infections," he says. "Every circumcision still has a preventive effect."

For people living in the epicentre of the HIV epidemic, that protective effect — however limited — is an attractive offering. Marvin, for one, is confident in his decision. "I think, for me," he says, "it's going to be a good thing." ■ [SEE EDITORIAL P.165](#)

**Catherine de Lange** is an editor with *Nature* in London. She travelled to Zambia on a fellowship from the International Reporting Project at Johns Hopkins University in Baltimore, Maryland.

1. Alcena, V. *NY State J. Med.* **86**, 446 (1986).
2. Cameron, D. W. *et al. Lancet* **334**, 403–407 (1989).
3. Auvert, B. *et al. PLoS Med.* **2**, e298 (2005).
4. Bailey, R. C. *et al. Lancet* **369**, 643–656 (2007).
5. Gray, R. H. *et al. Lancet* **369**, 657–666 (2007).
6. Njeuhmeli, E. *et al. PLoS Med.* **8**, e1001132 (2011).
7. Westercamp, M., Agot, K. E., Ndinya-Achola, J. & Bailey, R. C. *AIDS Care* **24**, 157–166 (2012).
8. Layer, E. H. *et al. PLoS ONE* **8**, e74391 (2013).
9. Auvert, B. *et al. PLoS Med.* **10**, e1001509 (2013).
10. Hewett, P. C. *et al. AIDS* **26**, 749–756 (2012).



# A race against RESISTANCE

*Several African nations could strike a major blow against malaria by sacrificing the efficacy of some older drugs. Can they make it work?*

BY AMY MAXMEN

**I**t is September in southeastern Mali, and Louka Coulibaly is standing in the shade of a squat, concrete building, giving instructions to a dozen men and women perched on a wobbly wooden bench. Coulibaly, a local medical supervisor, hands out nylon backpacks, each filled with bags of pills, plastic cups and a porcelain mortar and pestle that the women pause to admire. By noon, the men and women are packing up and heading back to their respective villages on foot, bicycle and motorcycle.

The following day, they and about 1,400 other health workers throughout the region will set up shop in public spaces: under the shade of mango trees, in one-room schools, at market stands and in district health centres. They will mix and mash the pills with the mortar and pestle, dissolve them in water in a cup, and hand the bitter dandelion-coloured liquid to about 164,000 children.

The effort is part of a broad campaign to prevent malaria by providing African children with drugs usually used to treat the disease. Nearly 1.2 million healthy children from parts of Mali, Togo, Chad, Niger, Nigeria and Senegal received these drugs during the rainy season — from around July to November — when malaria usually ravages the population.

The countries' governments are deploying this intervention — known as seasonal malaria chemoprevention, or SMC — with financial support from the United States, the United Nations and the medical aid organization Médecins sans Frontières (MSF), also called Doctors Without Borders. Next year, many plan to expand the campaigns, and other countries hope to launch their own, encouraged by

recommendations from the World Health Organization (WHO).

Preventive use of anti-malarial drugs is not new: tourists routinely swallow them when travelling. But public-health officials have long instructed people living in regions where the disease is endemic to refrain from taking drugs prophylactically, in part because of concerns that the parasite that causes malaria will develop resistance when many people take the medicine on a long-term basis.

That risk has not disappeared. In fact, scientists fully expect SMC to encourage widespread drug resistance. No one knows when, exactly, but it could happen within as few as five years. Until then, SMC has the power to prevent 8.8 million cases and 80,000 deaths each year if implemented in regions with high rates of seasonal malaria. That is considered a powerful enough benefit to justify losing the drugs.

"Life is a risk," says Coulibaly, a Malian hired by MSF to train local health workers. "And if you don't take risks, you don't win."

The project is designed to forestall drug resistance as long as possible, and to work in concert with mosquito nets and other preventive methods. Supporters hope that the combination will significantly suppress malaria, so that even if resistance eventually spreads, the caseload should be smaller and manageable with other treatments. But SMC will not be as successful if funding and infrastructure falter — and so far, programmes have had a shaky start. Still, advocates say that the challenges can be overcome.

**► NATURE.COM**  
For a video related to  
this story, see:  
[go.nature.com/gy6rv2](http://go.nature.com/gy6rv2)



"SMC is feasible," says Estrella Lasry, technical adviser for malaria chemoprevention at MSF. "But it requires organization, a series of steps and money to back it."

AMY MAXMEN

## WHEN INTENTIONS BACKFIRE

Previous attempts at large-scale malaria chemoprevention offer lessons on what not to do. In the 1950s, David Clyde, a malaria researcher with the British Colonial Medical Service, administered the drug pyrimethamine to villagers in Tanzania. At the time, pyrimethamine had a strong track record of clearing the parasite. But with any drug, there is a slim chance that some strains of parasite will be resistant and will survive to infect others — a chance that increases when many people take the medicine in an area where the parasites are abundant and circulate year-round.

Clyde's experiment drove this concept home: malaria rates dropped at first, but after five months, 37% of infections in the village no longer responded to the drug<sup>1</sup>. Eight years later, pyrimethamine resistance had spread: up to 40% of infections within 25 kilometres of the original intervention site were unresponsive.





In the Debela village in southeastern Mali, a mother gives her baby drugs meant to stave off malaria.

The 1960s brought more lessons — this time, when scientists tried adding the drug chloroquine to table salt. Clinical trials had shown<sup>2</sup> that the salt drastically lowered malaria rates. But when the tactic was scaled up and the salt was distributed to markets in Guyana and Brazil, people consumed only what met their tastes. Others opted for untreated salt when they could, because the chloroquine made their skin itch. As a result, many people carried sub-therapeutic levels of the drug — not enough to reduce the malaria burden, but enough to promote resistance. “The salt campaigns were a disaster,” says Christopher Plowe, a malariologist at the University of Maryland School of Medicine in Baltimore.

Governments and aid organizations mostly shelved chemoprevention programmes after that, but resistance continued to grow — albeit slowly — as people used drugs to treat malaria infections. Between 1960 and 2000, chloroquine resistance crept around the globe and the malaria death toll steadily rose. That trend started to reverse around 2005, after the widespread adoption of the drug artemisinin,

derived from Chinese sweet wormwood (*Artemisia annua*). Today, artemisinin-based drugs are the gold standard for treating malaria.

## SECOND CHANCE

Alassane Dicko, a malariologist at the University of Bamako in Mali, was a graduate student in Plowe’s laboratory in 2001, when he started to think seriously about reviving chemoprevention. As a child, Dicko had lost his older brother and his best friend to malaria. Later, as a medical student working in hospitals, he was distraught at the number of children he saw dying. “You really feel it,” he says. “If we want to do anything for this country in terms of health, we need to stop malaria first.”

Dicko suggested that older antimalarials might be repurposed for prevention in places where resistance to them is not yet widespread. By using drugs seasonally, only in uninfected children and in combination rather than alone, he hoped to avoid some of the mistakes of the past. With drug combinations, parasites need to acquire several mutations to survive. These mutations usually come at a cost to the parasite, so removing the selective pressure of the

drugs during the dry season would give parasites still sensitive to the treatment a chance to outcompete resistant ones.

Dicko proposed using a mixture of sulphadoxine and pyrimethamine called SP, which was known to be relatively safe over the long term. In 2002, his team treated 130 children with SP for two months in a placebo-controlled trial in Mali<sup>3</sup>. The treatment reduced malaria by 68%.

Other West African scientists followed the study. Among them was Badara Cissé, a Senegalese researcher then pursuing his doctorate with malariologist Brian Greenwood at the London School of Hygiene and Tropical Medicine. Greenwood had been considering chemoprevention since the 1980s, and he and Cissé immediately grasped the potential in Dicko’s approach. In 2004, they began a trial in Senegal to test three monthly doses of SP plus artesunate, an artemisinin derivative. Compared with the placebo group, nearly nine out of ten malaria cases were averted<sup>4</sup>.

With a US\$4.5-million grant from the Bill & Melinda Gates Foundation in 2008, Cissé and his colleagues launched an as-yet-unpublished, 3-year clinical trial to study SP with another drug, amodiaquine (to preserve the efficacy of artemisinin). They treated nearly 200,000 children under 10 years old and found that they had 83% fewer cases of malaria than controls, says Cissé. Smaller trials in other African nations reported similar findings.

These are impressive numbers, especially given how recalcitrant malaria has been to preventive measures. No vaccine has ever proved fully effective against the disease, for example. And the one that is closest to approval — RTS,S — has shown disappointing results in ongoing clinical trials, with less than a 50% reduction in cases (see *Nature* 502, 271–272; 2013).

## RESISTING THE CRITICS

SMC raised some concerns that slowed its adoption. Some health officials suggested that natural, partial immunity to the parasite — built up as a child survives multiple bouts of malaria — would be compromised. Others fretted about the potential side effects of taking the drugs regularly. But the loudest complaints were about losing the drugs to resistance.

In a cramped office in a makeshift building at the University of Dakar, Cissé explains how he was frustrated by the deliberations among public-health officials as malaria waged war on Senegal’s children. He slumps in a chair that seems much too small for him and asks, “Isn’t it selfish to sit in our offices with air conditioning, saying that we should save these drugs?” He recalls a single night, 20 years ago, when he watched five children die of malaria. There was nothing he could do to save them. “If this happened to you, you would not be debating about the fear of losing a drug,” he says.

In 2012, SMC finally won over most officials. The Cochrane Collaboration — an international group based in Melbourne,

Australia, that specializes in evidence assessment — analysed results from trials in Senegal, Mali, Burkina Faso, Ghana and Gambia, and concluded<sup>5</sup> that SMC could prevent more than three-quarters of malaria cases in places where the disease struck seasonally. In the trials, the signs of side effects, resistance and reduced immunity were all minimal. According to another report<sup>6</sup>, nearly 21 million children in these regions stood to gain from SMC each year. And prevention is cheaper than treatment. Each month, chemoprevention costs \$1.50 per child, which pales in comparison to the costs of travel and medical care for a child who falls ill. In November 2012, the WHO published SMC-implementation guidelines that enabled countries to apply for funds from international organizations<sup>7</sup>.

## SLOW START

Implementation has been a challenge, however. Mamadou Lamine Diouf, the drug-procurement manager for Senegal's National Malaria Control Program, says that the rollout there was supposed to reach nearly 600,000 children each month, starting in July and August. But he and the US agency footing the bill for the medicine had underestimated how much time it would take to get these older drugs manufactured anew and assessed by various organizations. By early November, health workers had managed to reach only 53,000 children. "We are learning by doing," says Diouf. "Now we know that if we don't master this long supply chain, nothing will be possible."

Drug delays set back chemoprevention pilots in northern Nigeria by a month. Togo's campaign did not start until September. Burkina Faso's project failed to launch when funds came up short. And the size of Mali's intended intervention dropped after a *coup d'état* and an invasion by al-Qaeda affiliates last year sent the nation into disarray.

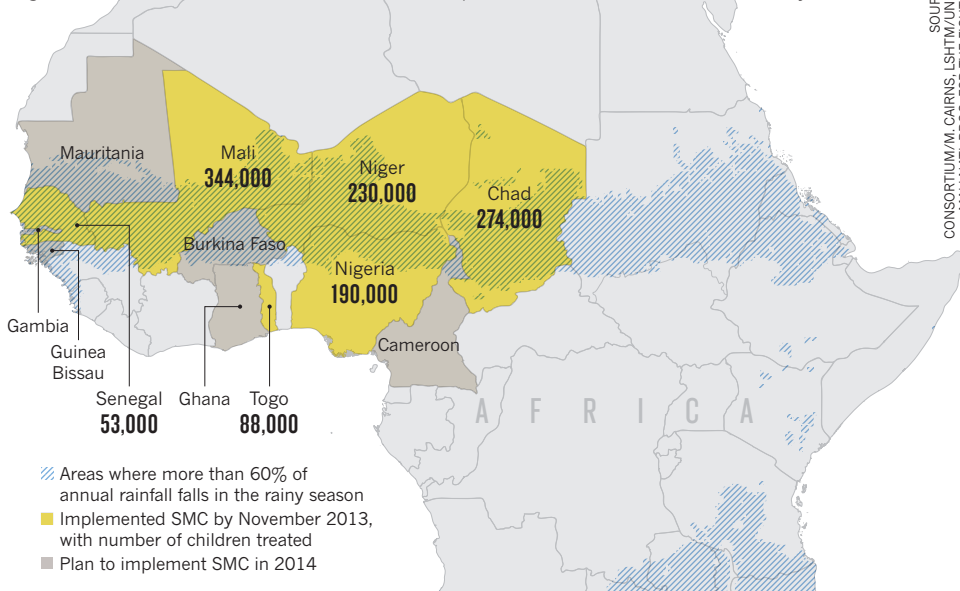
**"IF WE JUST ROLL THIS OUT WITHOUT SURVEILLANCE, WE RISK REPEATING ALL OF THE MISTAKES MADE IN THE PAST."**

Still, with the lessons learned, supporters say that they will be better prepared next year (see 'A million ounces of prevention'). In March, some countries plan to apply for funding from the Global Fund to Fight AIDS, Tuberculosis and Malaria. Scott Filler, a disease coordinator at the Global Fund, which is based in Geneva, Switzerland, says, "There are not many things that can prevent malaria in 75% of children, so we will fully support it when countries come to us."

As the programmes continue, researchers will keep watch to see if resistance to the drugs mounts. Randomly selected people who come to hospitals to be treated for malaria in Mali, Chad and Niger will have a spot of their blood

## A MILLION OUNCES OF PREVENTION

By November 2013, seasonal malaria chemoprevention (SMC) reached almost 1.2 million children in areas that receive at least 60% of their annual rainfall in the rainy season. If SMC were scaled up to cover all areas where it might be effective, it could reach 25 million children and prevent an estimated 80,000 deaths each year.



smear on filter paper, placed in a ziplock bag and shipped to a laboratory in Bamako, where Dicko and his colleagues will look for mutations associated with resistance to SP and amodiaquine. The University of Dakar will conduct similar tests.

For the campaigns to have a long-lasting effect, chemoprevention must work faster than the parasites acquire resistance. Supporters hope that the treatments will destroy most malaria parasites over the next several years, driving down infection rates and keeping them down even when resistance begins to spread.

Ramanan Laxminarayan, director of the Center for Disease Dynamics, Economics and Policy, a health-policy think tank in Washing-

village in the Koutiala region of Mali just after sunrise in September. Awa Damale, 25 years old and clad in an embroidered aqua dress and matching headscarf, arrives by donkey cart with her four children and two from another family. Five of the children swallow their medicine, but one of Damale's sons has felt ill this week. He tests positive for malaria and gets a referral to the nearest clinic. SMC is for prevention only.

The boy's illness may be a sign that the drugs he took last month are not 100% effective — or that he did not swallow all of the medicine — but his condition does not dampen Damale's enthusiasm. It is the first time this year that one of her children has had malaria. Before the intervention, she constantly juggled working on the farm with caring for sick children. She does not want to hear about the possibility of the programme drying up or the drugs losing potency years down the road. Most of her children are healthy now, and that is what matters most. ■

**SEE EDITORIAL P.165**

**Amy Maxmen** is a freelance science journalist in New York City. Travel for this story was paid for by a grant from the Pulitzer Center on Crisis Reporting in Washington DC.

1. Plowe, C. V. *Trans. R. Soc. Trop. Med. Hyg.* **103**, S11–S14 (2009).
2. Giglioli, G., Rutten, F. J. & Ramjattan, S. *Bull. World Health Org.* **36**, 283–301 (1967).
3. Dicko, A. *et al. Malar. J.* **7**, 123 (2008).
4. Cissé, B. *et al. Lancet* **367**, 659–667 (2006).
5. Meremikwu, M. M., Donegan, S., Sinclair, D., Esu, E. & Oringanje, C. *Cochrane Database Systematic Review* 2012 (<http://dx.doi.org/10.1002/14651858.CD003756.pub4> (2012).
6. Cairns, M. *et al. Nature Commun.* **3**, 881 (2012).
7. World Health Organization Seasonal malaria chemoprevention with sulfadoxine-pyrimethamine plus amodiaquine in children: A field guide (2012).



# COMMENT

**PSYCHOTROPICS** Two takes on the science, culture and politics of altered states of mind **p.194**

**EXHIBITION** Artists' riffs on synthetic biology engage the public in Dublin **p.196**

**TAXONOMY** Definitive guide to the world's birds goes online **p.197**

**METRICS** Is predicting impact akin to quantifying dreams? **p.198**



ANDREW BIRAU/REUTERS/CORBIS



In young children, the parasitic infection cryptosporidiosis is one of four leading causes of severe diarrhoea.

## Time to tackle cryptosporidiosis

The little-studied parasite *Cryptosporidium* is a major threat to infants. **Boris Striepen** calls on microbiologists and funders to give it more attention.

Earlier this year, a massive clinical and epidemiological study<sup>1</sup> involving 22,500 children from Africa and Asia revealed — unexpectedly — that the protozoan parasite *Cryptosporidium* is one of four pathogens responsible for the lion's share of severe diarrhoea in infants and toddlers. According to the World Health Organization, diarrhoea accounts for 10.5% of the nearly 8 million yearly deaths of children under five years of age<sup>2</sup>. (For comparison, malaria causes 7% of such deaths, and HIV/AIDS, just 2%.)

Vaccines and treatments are already available or fast being developed for three of the four pathogens identified: rotavirus, *Shigella* bacteria and enterotoxigenic *Escherichia coli* (see 'Child killer'). But for 'crypto', there is no fully effective drug treatment or vaccine, and the basic research tools and infrastructure needed to discover, evaluate and develop such interventions are mostly lacking.

The technical challenges of working on crypto in the laboratory have led to the perception that the pathogen is an intriguing yet

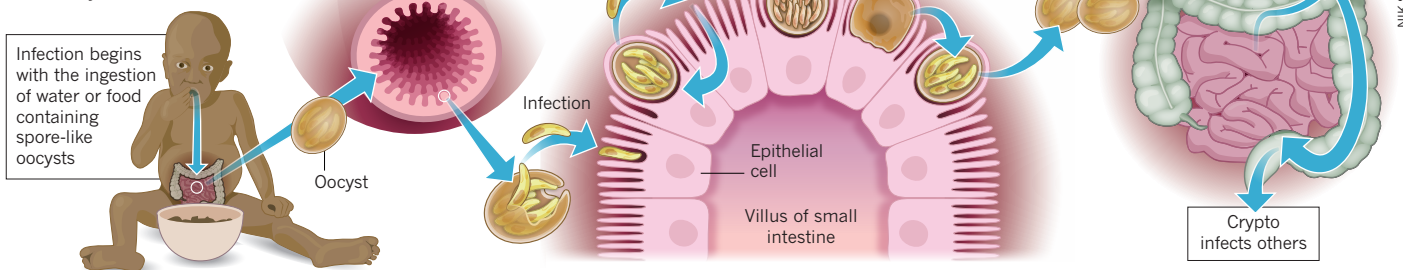
intractable problem. Crypto lab cultures last a few days at most, for instance, and some of the species that infect humans cannot be easily studied in standard model organisms such as mice. As a result, funders and biologists have tended to shy away from the parasite: a search on the biomedical research database PubMed suggests that in the past five years, about 20 times more articles have been published on malaria than on crypto.

I believe, however, that with the right tools, research approach and financial backing, the prevention of deaths caused



## EASY TARGETS

The parasite *Cryptosporidium* causes severe infections in young children and people with weak immune systems.



by crypto infections is a realistic goal. Furthermore, bringing crypto research into mainstream microbiology could provide fundamental mechanistic insights into intestinal infections in general.

## LINGERING PROBLEM

Infection with crypto occurs when people ingest spore-like oocysts, usually by drinking contaminated water (see 'Easy targets'). Parasites emerging from these spores invade the epithelial cells lining the small intestine, and cause the cells to form internal structures that house the parasite. These structures supply the parasites with nutrients and position them at the interface between the digestive tract and its contents. Infection generally results in one to two weeks of severe diarrhoea, but in people with HIV/AIDS or in malnourished children, the disease can linger and eventually kill.

Only one drug treatment for crypto is currently approved by the US Food and Drug Administration (FDA): nitazoxanide, marketed under the name Alinia. Clinical studies<sup>3</sup> in Egypt have shown that otherwise healthy children and adults suffering from diarrhoea caused by crypto are twice as likely to recover within a week when given the drug. But those people who are most vulnerable to the life-threatening disease need something better. In a study<sup>4</sup> in Zambia, diarrhoea symptoms lifted for fewer than 25% of malnourished children after treatment with nitazoxanide. Meanwhile, the drug seemed to have no effect in

children or adults whose immune systems are suppressed, for instance because of HIV infection.

Through horizontal gene transfer, crypto parasites have obtained various genes from the bacteria sharing their intestinal environment, some of which are known to be associated with bacterial disease. Crypto may also use products of bacterial metabolism to help it to grow and cause illness. By targeting the products of some of these genes, researchers could home in on treatments that disrupt pathways in the parasite but that do not affect human cells<sup>5</sup>.

Furthermore, crypto has a tiny genome, a simple life cycle requiring only one host and its reproduction is relatively straightforward. So it should be possible to cross crypto strains. Crossing strains that differ in the types of host they infect, for instance — such as healthy mice versus mice with a faulty immune system — could shed light on the genes and mechanisms causing disease.

Sequencing the genomes of two crypto species that infect humans<sup>6,7</sup> has revealed vulnerabilities stemming from the parasite's minimalist metabolism. Mammalian cells have multiple ways of producing certain metabolites; crypto often relies on just one, using components borrowed from its human host or bacteria. Drugs designed to target such pathways — interfering with enzymes that supply the building blocks of DNA and RNA, for example — show some promise.

Another option might be to repurpose drugs used to treat other illnesses<sup>8</sup>. A statin used to cut down cholesterol, for example, can interfere with the ability of crypto to get crucial lipids from its host cells. And efforts to find treatments could piggyback on existing large-scale programmes to identify malaria drugs — crypto is related to the *Plasmodium* parasites that cause malaria.

The main challenge to finding drug treatments for crypto, however, is the lack of tools for validating and testing potential targets or promising classes of compounds. Currently, assays are neither sufficiently robust nor scalable to exploit the high-throughput techniques typically used to screen hundreds of thousands of drug candidates.

Parasites that have been genetically engineered to synthesize proteins that produce light or colour, such as the green fluorescent protein from jellyfish, enable researchers to detect robust correlates of infection in a reproducible and cost-effective way for numerous infectious diseases, such as malaria and tuberculosis. Such technologies would be transformative for the study of crypto.

## STEMMING INFECTION

Even better than finding drugs to treat crypto effectively would be preventing the disease from taking hold in the first place.

Although crypto takes a massive toll on infants and toddlers, it is rarely detected in those older than two years, suggesting that children may develop natural immunity. This idea is supported by animal studies: in one, for instance, vaccines containing crypto protected calves and cattle<sup>9</sup>; in another, piglets previously infected with one crypto species later showed complete resistance to that same species, and showed mild symptoms on subsequent infection with a different species<sup>10</sup>.

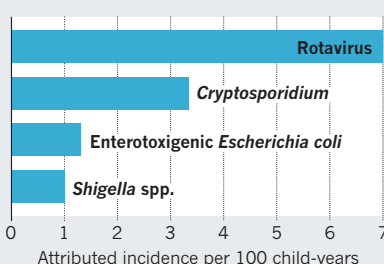
Small studies involving about 20 adult volunteers suggest that although infection cannot prevent subsequent disease, it can reduce its severity<sup>11</sup>. So developing vaccinations containing parasites that have been rendered incapable of causing disease, through irradiation or genetic engineering, is likely to be the most hopeful approach. Such vaccines have shown promise in preventing other parasitic infections such as malaria<sup>12</sup> and leishmaniasis.

Technology transfer from established to emerging research models can greatly aid innovation. For instance, applying knowledge and techniques associated with genetic modifications in fruitflies has revealed how mosquitoes smell and identify humans, offering avenues for the development of repellents and baits. For crypto, concepts and tools developed for the related parasites *Toxoplasma* and *Plasmodium*, including how to render them nonpathogenic, could prove useful.

Animal models that better mimic human infection and disease than current ones would allow researchers to study the impact

## CHILD KILLER

Cryptosporidiosis is the second biggest cause of diarrhoeal disease and death in infants.



of intestinal commensal bacteria on crypto infection. Calves and piglets, for instance, are natural hosts of the species of crypto that infects humans.

More money is key. The US National Institutes of Health (NIH) currently spends roughly US\$4.3 million each year on 12 crypto projects, compared to around \$300 million on more than 600 malaria projects.

Review panels for government agencies typically take a dim view of studies on non-model organisms that have a higher risk of failing than those on well-established systems, or that require large upfront investments in technology. Meanwhile, philanthropic organizations such as the Bill & Melinda Gates Foundation, based in Seattle, Washington, tend to focus on intervention rather than basic research. And industry may not see a market in a disease that mainly affects the poor.

Yet both the NIH and the Gates Foundation have already made substantial investments in projects designed to identify the causes of early childhood diarrhoea. Now these organizations have a responsibility to act on their discoveries. Furthermore, crypto kills young animals too, so the possibility of preventing it in livestock should stimulate industry interest, resources and expertise.

As well as potentially saving hundreds of thousands of young lives and reducing untold suffering for infants, children and parents, I am confident that bringing crypto to the forefront of microbiology research will have wide-ranging benefits. It is likely to be an excellent model system for understanding the complex interaction between pathogens, host nutrition, inflammation and the intestinal microbiome. ■

**Boris Striepen** is professor and Georgia Research Alliance distinguished investigator at the Center for Tropical and Emerging Global Diseases, University of Georgia, Athens, Georgia 30602, USA.  
e-mail: [striepen@uga.edu](mailto:striepen@uga.edu)

1. Kotloff, K. L. *et al.* *Lancet* **382**, 209–222 (2013).
2. Liu, L. *et al.* *Lancet* **379**, 2151–2161 (2012).
3. Rossignol, J.-F. A., Ayoub, A. & Ayers, M. S. *J. Infect. Dis.* **184**, 103–106 (2001).
4. Amadi, B. *et al.* *Lancet* **360**, 1375–1380 (2002).
5. Umejiego, N. N. *et al.* *Chem. Biol.* **15**, 70–77 (2008).
6. Abrahamsen, M. S. *et al.* *Science* **304**, 441–445 (2004).
7. Xu, P. *et al.* *Nature* **431**, 1107–1112 (2004).
8. Bessoff, K., Sateriale, A., Lee, K. K. & Huston, C. D. *Antimicrob. Agents Chemother.* **57**, 1804–1814 (2013).
9. Jenkins, M., Higgins, J., Kniel, K., Trout, J. & Fayer, R. *J. Parasitol.* **90**, 1178–1180 (2004).
10. Sheoran, A., Wiffin, A., Widmer, G., Singh, P. & Tzipori, S. *J. Infect. Dis.* **205**, 1019–1023 (2012).
11. Okhuysen, P. C., Chappell, C. L., Sterling, C. R., Jakubowski, W. & DuPont, H. L. *Infect. Immun.* **66**, 441–443 (1998).
12. Seder, R. A. *et al.* *Science* **341**, 1359–1365 (2013).



# Ecology must evolve

Tackling global problems requires a fresh approach, argues **Georgina Mace**, as the British Ecological Society celebrates its centenary.

Climate change, the threat of pandemics, population growth, food security and the loss of biodiversity and ecosystem services demand a new kind of ecology — one that focuses on how whole communities of organisms, at the scale of landscapes or catchments, interact with people and the physical environment.

The advances in ecology in the past century have hugely improved our understanding of species interactions, such as those between hosts and parasites or between predators and prey, as well as population dynamics, food-web dynamics and how organisms adapt to their local environments. Such gains have come mostly from a combination of theory and modelling, and carefully designed long-term laboratory or field experiments<sup>1</sup> in places as diverse as the Serengeti Wildlife Research Centre in Tanzania and the University of Oxford's Wytham Woods site, UK. Indeed, historically, careers in ecology have tended to revolve around the

'ownership' and analysis of a personal study system or a painstakingly curated data set.

In part because of this history, there are few general theories for how multiple species respond to perturbations, such as disease or shifting weather patterns, at the community level<sup>2,3</sup>. This is a major problem for what is becoming known as global-change science.

The British Ecological Society, founded in 1913 "to promote and foster the study of Ecology in its widest sense", has been key in disseminating knowledge among ecologists over the past few decades, by publishing leading journals and organizing academic meetings, and through its increasing presence in education and policy. As its outgoing president, however, I worry that in today's world, there are more pressing demands. And, as for other scientific societies that are reliant on revenue from academic journals, the shift to electronic and open-access publishing could weaken its business model.

A new role for the BES and similar ►

of intestinal commensal bacteria on crypto infection. Calves and piglets, for instance, are natural hosts of the species of crypto that infects humans.

More money is key. The US National Institutes of Health (NIH) currently spends roughly US\$4.3 million each year on 12 crypto projects, compared to around \$300 million on more than 600 malaria projects.

Review panels for government agencies typically take a dim view of studies on non-model organisms that have a higher risk of failing than those on well-established systems, or that require large upfront investments in technology. Meanwhile, philanthropic organizations such as the Bill & Melinda Gates Foundation, based in Seattle, Washington, tend to focus on intervention rather than basic research. And industry may not see a market in a disease that mainly affects the poor.

Yet both the NIH and the Gates Foundation have already made substantial investments in projects designed to identify the causes of early childhood diarrhoea. Now these organizations have a responsibility to act on their discoveries. Furthermore, crypto kills young animals too, so the possibility of preventing it in livestock should stimulate industry interest, resources and expertise.

As well as potentially saving hundreds of thousands of young lives and reducing untold suffering for infants, children and parents, I am confident that bringing crypto to the forefront of microbiology research will have wide-ranging benefits. It is likely to be an excellent model system for understanding the complex interaction between pathogens, host nutrition, inflammation and the intestinal microbiome. ■

**Boris Striepen** is professor and Georgia Research Alliance distinguished investigator at the Center for Tropical and Emerging Global Diseases, University of Georgia, Athens, Georgia 30602, USA.  
e-mail: [striepen@uga.edu](mailto:striepen@uga.edu)

1. Kotloff, K. L. *et al. Lancet* **382**, 209–222 (2013).
2. Liu, L. *et al. Lancet* **379**, 2151–2161 (2012).
3. Rossignol, J.-F. A., Ayoub, A. & Ayers, M. S. *J. Infect. Dis.* **184**, 103–106 (2001).
4. Amadi, B. *et al. Lancet* **360**, 1375–1380 (2002).
5. Umejiego, N. N. *et al. Chem. Biol.* **15**, 70–77 (2008).
6. Abrahamsen, M. S. *et al. Science* **304**, 441–445 (2004).
7. Xu, P. *et al. Nature* **431**, 1107–1112 (2004).
8. Bessoff, K., Sateriale, A., Lee, K. K. & Huston, C. D. *Antimicrob. Agents Chemother.* **57**, 1804–1814 (2013).
9. Jenkins, M., Higgins, J., Kniel, K., Trout, J. & Fayer, R. *J. Parasitol.* **90**, 1178–1180 (2004).
10. Sheoran, A., Wiffin, A., Widmer, G., Singh, P. & Tzipori, S. *J. Infect. Dis.* **205**, 1019–1023 (2012).
11. Okhuysen, P. C., Chappell, C. L., Sterling, C. R., Jakubowski, W. & DuPont, H. L. *Infect. Immun.* **66**, 441–443 (1998).
12. Seder, R. A. *et al. Science* **341**, 1359–1365 (2013).



# Ecology must evolve

Tackling global problems requires a fresh approach, argues **Georgina Mace**, as the British Ecological Society celebrates its centenary.

Climate change, the threat of pandemics, population growth, food security and the loss of biodiversity and ecosystem services demand a new kind of ecology — one that focuses on how whole communities of organisms, at the scale of landscapes or catchments, interact with people and the physical environment.

The advances in ecology in the past century have hugely improved our understanding of species interactions, such as those between hosts and parasites or between predators and prey, as well as population dynamics, food-web dynamics and how organisms adapt to their local environments. Such gains have come mostly from a combination of theory and modelling, and carefully designed long-term laboratory or field experiments<sup>1</sup> in places as diverse as the Serengeti Wildlife Research Centre in Tanzania and the University of Oxford's Wytham Woods site, UK. Indeed, historically, careers in ecology have tended to revolve around the

'ownership' and analysis of a personal study system or a painstakingly curated data set.

In part because of this history, there are few general theories for how multiple species respond to perturbations, such as disease or shifting weather patterns, at the community level<sup>2,3</sup>. This is a major problem for what is becoming known as global-change science.

The British Ecological Society, founded in 1913 "to promote and foster the study of Ecology in its widest sense", has been key in disseminating knowledge among ecologists over the past few decades, by publishing leading journals and organizing academic meetings, and through its increasing presence in education and policy. As its outgoing president, however, I worry that in today's world, there are more pressing demands. And, as for other scientific societies that are reliant on revenue from academic journals, the shift to electronic and open-access publishing could weaken its business model.

A new role for the BES and similar ►



► societies could be to help drive the kind of ecological understanding that would guide and inform efforts to address today's greatest global challenges.

### EXPANDING HORIZONS

No matter how intriguing the details and well understood the dynamics and processes, the results of small-scale ecological studies are difficult to tie into models simulating Earth systems such as the climate, ocean circulation or the water cycle. Yet as long as researchers model such systems without taking ecological processes into account, they risk missing significant feedbacks, thresholds and constraints.

Insights from long-term studies, for instance of red deer on the Scottish isle of Rum, or of various species inhabiting Yellowstone National Park in Wyoming, must be tied into global-change science. In fact, a new kind of ecology is needed that is predicated on scaling up efforts, data sharing and collaboration<sup>4–8</sup>. This would link up with related disciplines in the environmental, social and physical sciences, and focus much more on meta-analyses and synthesis. We need to identify general trends, such as the effect of a change in temperature on animal dispersal and the knock-on consequences for ecosystems, and draw general conclusions, for instance on how the practices of local fisherman affect the health of fisheries. A few scaled-up ecological models do exist, notably the emerging Madingley model, which is an attempt to simulate global ecosystems<sup>9</sup>. But many more are needed.

The European Research Council, UK research councils and the US National Science Foundation are being increasingly sparing with grants to single investigators. Instead, the trend is to award larger sums (US\$2 million to \$10 million) to multi-skilled teams of researchers tackling broad environmental questions. Presumably, it is easier and cheaper to give large sums to a few people than smaller sums to many different recipients. But this preferential funding of large-scale programmes can penalize people who are still in the early stages of their careers, because researchers are likely to be invited into such teams only after they have established themselves. It also threatens the funding of long-term ecological studies, some of which are already proving invaluable to better understanding the ecological consequences of human-induced alterations, such as climate change.

Societies such as the BES, the Ecological Society of America, the Society for Conservation Biology and related scientific societies for meteorology, geology and geography could help to forge a new ecology that would be both inclusive and innovative. Goals



would be shaped by scientists, policy-makers and users of the resulting science, rather than by recent publishing trends — which is often what drives allocations from funders at present.

### HARNESS COLLABORATION

Like most scientific societies, the BES relies mainly on revenue from owning and publishing academic journals, such as the *Journal of Ecology* and the *Journal of Animal Ecology*. These days, however, open-access journals, and publishers offering libraries and institutions bulk deals for access to many journals, have made individual subscriptions less necessary. In many societies, membership — which for young researchers is now more about access to conferences, researchers with shared interests, and links to industry and policy than to journals — has been falling in recent years.

I believe that the mass of expertise and breadth of focus from basic science to policy found among members of the BES and other societies, as well as among their contacts and collaborators, could be used to help identify the grand challenges for ecology and environmental-change science, and to design and develop research priorities.

To harness this knowledge, society staff need to bring the right people together, for instance, through focused meetings, and by funding placements for scientists in government and in industry. Research councils and industry may be willing to fund such activities on the grounds that the resulting collaborations would help to address their own problems and questions. Schemes run by the BES, such as those placing PhD students in the UK Department for Environment, Food and Rural Affairs, are hugely beneficial to everyone involved. Such efforts need scaling up.

Furthermore, huge gaps in the science and supporting technologies need to be closed in many areas, from remote sensing and ecological modelling to 'environmental genomics' — the study of genome-level

interactions with the environment. Citizen-science innovations such as iSpot, an online forum for sharing skills and knowledge for species identification, illustrate the potential both to improve people's knowledge of plants and animals, and to monitor changes in species' phenology, abundance and distribution at low cost.

Scientific societies could help to drive the expansion of such data-gathering technologies in which the interests of users, producers and researchers align. For instance, iSpot, currently a UK-based initiative, would have enormous value globally. Societies could also facilitate the bringing together of related, but operationally disparate, projects to achieve some common goal.

Often missing from large-scale studies of the environment and our interactions with it, is the crucial first step of framing, testing and reframing the conceptual basis for the work. From my experience as a coordinating lead author in both the Millennium Ecosystem Assessment (MEA) and the UK National Ecosystem Assessment, this is key to achieving interdisciplinary understanding of terms and concepts and to articulating the most important questions. For example, in the case of the MEA, identifying outcomes related to human well-being — such as food security or freedom from health problems — and the main drivers of ecosystem change related to ecosystem services, allowed experts from diverse disciplines to undertake the large, complex assessment of the world's ecosystems.

By mediating this kind of conceptual framing for global-scale problems, numerous scientific societies would have a way to support their disciplines and keep them relevant to society — now and in the future. ■

**Georgina Mace** was president of the British Ecological Society from 2011 to 2013. She is professor of biodiversity and ecosystems, and director of the Centre for Biodiversity and Environment Research at University College London, UK.

e-mail: g.mace@ucl.ac.uk

1. British Ecological Society. *100 Influential Papers* (BES, 2013).
2. Ricklefs, R. E. *Am. Nat.* **172**, 741–750 (2008).
3. Lawton, J. H. *Oikos* **84**, 177–192 (1999).
4. Sutherland, W. J. *et al. J. Ecol.* **101**, 58–67 (2013).
5. Carpenter, S. R. *et al. BioScience* **59**, 699–701 (2009).
6. Hampton, S. E. & Parker, J. N. *BioScience* **61**, 900–910 (2011).
7. Michener, W. K. & Jones, M. B. *Trends Ecol. Evol.* **27**, 85–93 (2012).
8. Reichman, O. J., Jones, M. B. & Schildhauer, M. P. *Science* **331**, 703–705 (2011).
9. Purves, D. *et al. Nature* **493**, 295–297 (2013).



A poster from the 1920s reflects the era's fashion for drug-taking.

## HUMAN BEHAVIOUR

# Bending minds, building medicine

**Andrew Robinson** assesses two studies on the culture, biology and chemistry of psychotropic substances.

From drinking espresso to shooting heroin, the human impulse towards altered mental states is universal. Now two books explore the history of individual and social attitudes to psychotropic drugs, and the science of pharmacology.

In *Demons*, science historian Virginia Berridge focuses on drug cultures, primarily in Britain and to a lesser extent in the United States, Europe and Asia. She chronicles the consumption of opium, alcohol, tobacco, cocaine and cannabis, from the nineteenth

century onwards, and examines attempts at official control. Berridge points out that a drug's acceptability to mainstream society fluctuates more owing to social and cultural trends than to medical knowledge. The periodic reclassification of drugs is "not a 'rational' process", she concludes.

*Drugged*, by pharmacologist Richard Miller, offers the biochemical research absent from *Demons*. We visit university and industrial laboratories, learn about medical trials and clinical applications, and explore

## **Demons: Our Changing Attitudes to Alcohol, Tobacco, and Drugs**

VIRGINIA BERRIDGE

Oxford University Press: 2013.

## **Drugged: The Science and Culture Behind Psychotropic Drugs**

RICHARD J. MILLER

Oxford University Press: 2013.

the structures of organic molecules and their functions. Despite copious diagrams, much of the technical detail will be impenetrable to the generalist, unlike the accessible *Demons*. However, Miller casts his net wider than does Berridge, covering everything from tea and coffee to LSD (lysergic acid diethylamide) and Prozac (fluoxetine hydrochloride), and not limiting himself geographically. He also makes lengthy and lively excursions into culture.

Opium gets the most attention in both books. The latex from the opium poppy *Papaver somniferum* (from which opium and opiates such as morphine come) has been cultivated and used medically and recreationally for millennia, and began to seep into the West in the sixteenth century. As both Berridge and Miller discuss, the opium wars of the mid-nineteenth century, which pitted China against the French and British, played a key part in international regulation of drugs in the twentieth century. Opium was ubiquitous in Victorian Britain, almost like aspirin today, notes Berridge. Prized as a painkiller and stimulant, it was prescribed for everything from heart disease to gallstones, and taken as pills, lozenges, powder and even enemas. In the marshy Fens of eastern England, opium-spiked beer soothed fevers, chills and rheumatism. Prime Minister William Gladstone is said to have taken opium in coffee before big speeches.

By mid-century, the tincture laudanum was the most widely used form of opium. Medicinal use often led to addiction, as with early adopters such as Thomas De Quincey, author of the autobiographical *Confessions of an English Opium-Eater* (1821). Wilkie Collins, whose detective novel *The Moonstone* (1868) featured opium addiction, became so dependent on laudanum that he eventually took it in quantities large enough to kill an unhabituated person, and suffered paranoid delusions. Even Arthur Conan Doyle's fictional detective Sherlock Holmes, created in the 1880s, injected himself with morphine and cocaine for "mental exaltation", although in the final stories Holmes has been

Laudanum (opium tincture) was a Victorian panacea.



Laudanum (opium tincture) was a Victorian panacea.

SCIENCE MUSEUM, LONDON/WELLCOME IMAGES



weaned off the habit, presumably in deference to growing social disapproval.

Opium, alcohol and cocaine were also active ingredients in many US patent medicines in the nineteenth century. By 1900, the modern concept of addiction was beginning to emerge and the tide began to turn against cocaine and opium. In 1913, the last officially traded chest of opium was shipped from British India to China. In 1914, the US government passed the Harrison Narcotics Tax Act to limit the use of all opiates and coca products to medicine. In the 1920s, the influential International Opium Convention heralded the start of global drug control.

Cannabis (hemp) features more prominently in *Drugged* than in *Demons*. Although the drug was available in the United States long before, in the 1930s Mexican migrant labourers introduced North America to the practice of marijuana smoking. Marijuana was adopted first by black jazz musicians, and then, from the 1950s, by white society. Its demonization by the US government began with the federal Marihuana Tax Act of 1937, which made it prohibitively expensive for any purpose, including medicinal use. As Miller powerfully describes, the legislation was the result of racist scare tactics against Mexicans and black people by Harry Anslinger, the commissioner of the Federal Bureau of Narcotics from 1930 to 1962. He was supported by leaders of the synthetics industry, who wanted to end nascent US hemp production.

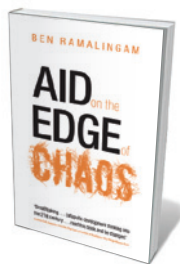
As for alcohol, the United States' nationwide prohibition from 1920 was repealed in 1933, during the Great Depression. The federal government needed the tax revenue. Moreover, attitudes had changed: prohibition had not solved industrial issues such as absenteeism, yet it had encouraged organized crime, political corruption and a trade in poisonous illicit alcohol. However, as Berridge stresses, prohibition also significantly reduced US drinking and alcohol-related disease, and might have worked had it been less restrictive. It would have failed in Britain, she argues, because of the complex British attitude to drunkenness. Norman Kerr, a physician and founder of the Society for the Study and Cure of Inebriety in 1884, asked whether inebriation was "a sin, a crime, a vice or a disease". This is a question that many are still attempting to answer in binge-drinking Britain.

However the use and abuse of specific drugs may shift over time and space, the need to bend the mind seems tightly tied to human culture. Berridge and Miller ably unpick the threads of that universal urge. ■

**Andrew Robinson** is the editor of *The Scientists: An Epic of Discovery*.  
e-mail: [andrew.robinson33@virgin.net](mailto:andrew.robinson33@virgin.net)

► **NATURE.COM**  
W. F. Bynum reviews  
an exhibition on  
mind-altering drugs:  
[go.nature.com/uciqym](http://go.nature.com/uciqym)

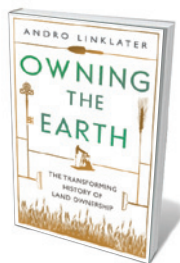
## Books in brief



### **Aid on the Edge of Chaos: Rethinking International Cooperation in a Complex World**

Ben Ramalingam OXFORD UNIVERSITY PRESS (2013)

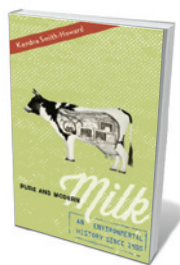
Many see international development aid as in thrall to linear, mechanized thinking, and champion approaches in which local people solve their own challenges with intelligently tailored backing. Complex adaptive systems thinking, argues Ben Ramalingam, offers a scientific model for that path. Awareness of real-world dynamics would, he avers, promote fine targeting of aid efforts and foster a new aid paradigm: an "open innovation network, catalysing and leveraging change in countries around the world".



### **Owning the Earth: The Transforming History of Land Ownership**

Andro Linklater BLOOMSBURY (2013)

Historian Andro Linklater hinges his layered chronicle of the past five centuries on land ownership, a growing concern on a crowded Earth. Arguing that land claims work only when yoked to public interest, he shows they were also rare up to 1800: before then, much of Europe was jointly 'owned' by peasants and royalty, for instance. The tide had begun to turn centuries before, however, when English freeholders gained unprecedented power. A brilliant treatise on the social and political shifts in our journey from indigenous 'commons' to the subprime-mortgage catastrophe of 2007 and beyond.



### **Pure and Modern Milk: An Environmental History since 1900**

Kendra Smith-Howard OXFORD UNIVERSITY PRESS (2013)

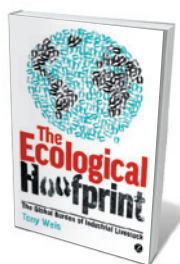
From cream and cheese to milk bottled, dried and lurking in everything from cake to glue, 'dairy' is ubiquitous. Yet getting the highly perishable, machine-pumped product of lactating cows to consumers has been a hugely complex technological, cultural and political saga. Kendra Smith-Howard deftly traces that trajectory in the United States since 1900. Initially rife with bacteria such as *Salmonella typhi*, milk became today's 'natural' staple through scientific manipulations — in bovine breeding, bacteriology, processing technology, and refrigerated bulk storage and transport.



### **In Meat We Trust: An Unexpected History of Carnivore America**

Maureen Ogle HOUGHTON MIFFLIN HARCOURT (2013)

In 2011, Maureen Ogle informs us, the United States produced 42 billion kilograms of pork, poultry and beef. The downsides are considerable: US livestock production and the meat industry have drawn fire for everything from pollution to overuse of pharmaceuticals. Ogle delivers a well-researched history of the US meat habit in the past 200 years. Although she decries twenty-first-century feedlots and "manure lagoons", Ogle argues that the culprits are not solely industrial agriculture and government laxity: US consumers' sense of entitlement to cheap meat is the real issue.



### **The Ecological Hoofprint: The Global Burden of Industrial Livestock**

Tony Weis ZED BOOKS (2013)

This cogent analysis of rising global meat consumption puts the blame firmly on industrial livestock production — a meat 'monoculture' that has quadrupled in the past 50 years. Agrarian political economist Tony Weis calculates that the cost to the environment and animal welfare is unsustainable, and the effect on social equity negative: livestock consume a third of all grain, while one person in seven remains malnourished or hungry. **Barbara Kiser**





In *Stranger Visions*, a work in *Grow Your Own*, Heather Dewey-Hagborg creates portraits from found DNA.

#### SYNTHETIC BIOLOGY

## The second creation

**Anthony King** finds himself on some of the wilder shores of synthetic biology at Dublin's Science Gallery.

Peering into a fridge, I am invited to sniff a cream cheese labelled “Christina”. It is crafted from milk and starter cultures sampled from someone’s skin. There are three more, each with a unique set of microbes from a different person: cheesy doppelgängers that will never, in our pasteurized world, reach the table.

*Selfmade* is the work of synthetic biologist Christina Agapakis and smell provocateur Sissel Tolaas, who are making the human microbiome tangible and highlighting a potential conflict between our antibacterial present and a future filled with engineered bacteria. Agapakis and Tolaas are members of Synthetic Aesthetics, a project run by the University of Edinburgh, UK, and Stanford University in California, to research the borders between art and synthetic biology. Theirs is one exhibit in the Dublin Science Gallery’s *Grow Your Own ... Life After Nature* show,

**Grow Your Own ... Life After Nature**  
Science Gallery, Trinity College, Dublin.  
Until 19 January 2014.

which explores genetic engineering. The show, which is split into sections labelled ‘Life’, ‘Society’ and ‘Machines’, offers a brew of utopian and dystopian views. There are also chances for visitors to ‘grow their own’ life form in a biolab and, as usual in this venue, there is an emphasis on citizen engagement with science.

In the Life section, I ponder *Escherichia coli* that has been modified to smell of banana, and then an exhibit exploring the idea that a woman could carry the fetus of a rare dolphin. In Society, upstairs, I discover how to ‘biohack’ the first commercially available genetically modified flower — the blue Moondust carnation — by using basic plant-culture methods to re-root the cut flowers and plant them to rejoin the carnation gene

pool. This do-it-yourself project raises questions about the ownership of life, including the moral question of whether the hacked plants should be released.

Artist David Benqué gives the heated debate on genetically modified organisms a novel twist. In his installation *The New Weathermen*, fictional environmental activists use madcap synthetic-biology schemes to fight corporate biotechnology interests. The exhibit takes the form of their scaled-down rigs, such as a “self-inflating antipathogenic membrane pump” and an “autonomous seed disperser”, with which Benqué’s anarchists spray golf-course grass modified for herbicide tolerance to spread their own open-source super-grass.

In the Machines section sits a piece by design futurist Natsai Audrey Chieza. Chieza collaborates with microbiologist John Ward at University College London to find soil bacteria that produce pigments suitable for dyeing textiles. In *Faber Futures* — a reference to the Latin for ‘to create’ — Petri dishes are filled with *Streptomyces* bacteria from soils in which different herbs were growing. Chieza removed specific colonies displaying various colours to dye silk scarves: bacteria isolated from around tarragon yielded blood-red hues; spearmint, deep purple; and oregano, light blue to navy. “Sometimes the most beautiful colours emerge if a contaminant comes in,” says Chieza.

Show curator and exhibitor Alexandra Daisy Ginsberg’s *Designing for the Sixth Extinction* in the Society section is a naturalistic ‘photo’ of a verdant forest inhabited by strange, digitally inserted creatures — such as a roving seed-dispersal device to increase plant biodiversity. Patented life forms released into the wild to assist endangered species, they are constructed using an expanded DNA code of six bases rather than the usual four. This produces non-biodegradable proteins, which are more robust. “It sets up a future where we could preserve and maintain nature using synthetic biology,” says Ginsberg.

Most people know little about synthetic biology, despite the best intentions of leaders in this field such as Craig Venter and George Church. “The foundations are being laid now,” Ginsberg says. “That is why it’s so important that people understand what it is and the many different directions it could go in.” *Grow Your Own* urges us to imagine many possible engineered futures — positive and negative, created with or without our consent — and challenges us to engage with the research by joining in. In January, visitors will even be able to use DNA to nudge *E. coli* into fluorescing in the exhibition biolab. ■

**Anthony King** is a writer based in Dublin.  
e-mail: anthonyking@gmail.com

# World wide wings

Stuart Pimm joyously clicks his way through an ornithological classic as it goes online.

I confess to knowing how many species of bird I have seen (3,763) and that just over 200 people report having seen more. The eccentricity of ornithologists may be exceptional, but an interest in birds is not. Bird watchers are abundant (the Royal Society for the Protection of Birds is Britain's biggest nature-conservation charity) and ubiquitous (there are bird lists for the North and South Poles, and surely everywhere in between). So the combined knowledge of birdwatchers is prodigious. In the *Handbook of the Birds of the World Alive* ([www.hbw.com](http://www.hbw.com)), edited by a team led by Josep del Hoyo, that knowledge now has an online home, with a basic annual subscription available for €29.95 (US\$41).

The online handbook defines what we know about the biogeography, behaviour and ecology of the almost 10,000 extant avian species at a time of unprecedented environmental change. It allows the compilation of instant, custom-made field guides. Its facility for species comparisons will accelerate comparative studies, and embed user-supplied data will open up new research opportunities.

The first volume of *Handbook of the Birds of the World* was published in 1992 by Lynx Edicions of Barcelona, Spain. Some 15 million words, 1,000 colour plates and 100,000 references later, the seventeenth and

## Handbook of the Birds of the World Alive

EDITED BY JOSEP DEL HOYO ET AL.  
Lynx: 2013.

final volume arrived this year. These gorgeous books carry the limitations of print, of course — for instance, they demand the

tedious perusal of indexes to find information, and provide no easy connection to the materials quoted.

The online version makes obvious questions easy. So, for 'what?' I click on a thumbnail of gulls and get a plate, scaled by size, of the 51 species. For 'where?' I filter the search by Chile and Argentina and find 14 species I might see on an upcoming trip. I could have added terns, other bird families or other countries. The 'compare' feature stacks similar species, such as Belcher's and Olrog's gulls, side by side. The taxonomy section tells me that taxonomists once considered these two birds to be the same species. Maps show that the former lives on Pacific coasts and the latter on Atlantic ones; text compares descriptions, habitats and natural history.

The International Union for Conservation of Nature's Red List gives Olrog's gull 'vulnerable' status. Using an embedded link, I access the curated Internet Bird Collection: 9,312 bird species feature in a collection that includes almost 100,000 photos and 80,000 videos. Ten of the items for Olrog's gull are connected to a Google Earth map. I shall soon sail past one of these locations. The prospect of a sighting is unreasonably exciting.

For many species, the best way to find and identify birds is through their call, so I click the link for Xeno-canto, the online database of bird calls and songs. It provides both a recording and a sonogram. Another

A roseate spoonbill (*Plateala ajaja*) soars through the sky in Florida, USA.

link takes me, via Cornell University's Lab of Ornithology, to eBird — a massive, global, crowd-sourced compilation of observations. A link to BirdLife International tells me why Olrog's gull is threatened. Active links to original sources are included, too. I can add my comments online and point to egregious failures to quote every last one of my published bird papers, however obscure. And Google's translation technology means that the whole resource can be read in 71 languages.

The march of online and crowd-sourced data is changing research completely and goes far beyond birds. For example, through smartphone apps such as iNaturalist, amateurs and professionals alike are sharing information on the species they are passionate about — be they frogs, seahorses or wild pigs — from Aiguabarreig Segre-Cinca in Catalonia, Spain, to Xishuangbanna in Yunnan, China. Through these, we are rapidly learning precisely where species are and so can quickly pinpoint where habitat loss and climate disruption may move them — and where extinctions happen.

The handbook's users can check off the species they have seen in their lifetime. Sightings are reported for 14 of the 20 species that have recently become extinct or extinct in the wild, meaning that extinctions are current, not historical, events. The handbook lists 176 species as 'critically endangered'. When, and how often, birders last counted these provides the grist for the statistical mill that predicts how likely it is that a species still survives, and so can be saved (see S. Pimm *Nature* 426, 235–236; 2003). Out of eccentricity comes insight and, perhaps, even conservation action. ■

Stuart Pimm is professor of conservation at the Nicholas School of the Environment, Duke University, Durham, North Carolina, USA, and author of *The World According to Pimm: A Scientist Audits the Earth*.  
e-mail: [stuartpimm@me.com](mailto:stuartpimm@me.com)



4. Goldie's Bird-of-paradise  
(*Paradisaea decora*) ♂



5. Red Bird-of-paradise  
(*Paradisaea rubra*) ♂



6. Emperor Bird-of-paradise  
(*Paradisaea guilielmi*) ♂

The online *Handbook of the Birds of the World Alive* allows a comparison of similar species.



# Correspondence

## Impact: akin to quantifying dreams

In our haste to measure everything in order to wring out evidence that non-specialists can understand and to secure funding, we forget that predicting the impact of research is akin to quantifying dreams (*Nature* **502**, 271; 2013). There are no short cuts for proper research assessment.

The impact of research on society is a composite of many strands of work, usually by different scientists and engineers, which — often serendipitously — culminate years later in changing some aspect of our lives. But attempting to disentangle those threads is a hopeless task.

There is probably little prognostic value in counting research-paper downloads, for example. In fact, such metrics are but surrogates of real research impact and can generate goals of their own. They encourage 'gaming', or manipulation of data to artificially improve metrics.

When used over time within institutions, metrics can be useful guides — we all need external measures of some sort. It is when they are used as a form of currency in their own right that we get into trouble.

**Jim Woodgett** *Lunenfeld-Tanenbaum Research Institute, Toronto, Canada.*  
woodgett@lunenfeld.ca

## Impact: take peer review into account

Recognition of a researcher's impact should include a measure of his or her contribution as a peer reviewer in maintaining high scientific standards in research papers and grant applications (*Nature* **502**, 287; 2013).

It should be feasible to create a reviewers' equivalent of the *h*-index — which measures the impact of research output in terms of its quantity and popularity. This might indicate, for example, the number and

impact of journals and papers for which the reviewer has acted as a referee. Publishers releasing these figures would need to protect the blind or double-blind review process.

They could do this by using a central repository to assign these metrics to particular researchers through identifiers such as the Open Researcher and Contributor ID scheme (see *Nature* **485**, 564; 2012).

As well as securing the reputation of reviewers, such a measure would benefit journals and publishers by encouraging more scientists to undertake refereeing.

**E. Tobias Krause** *Bielefeld University, Germany.*  
tobias.krause@uni-bielefeld.de

## Impact: China needs to review its metrics

Impact metrics are creating their own problems in China (see *Nature* **502**, 271; 2013).

By recognizing only the first and corresponding authors on a paper, for example, China's metrics are indirectly discouraging participation in research consortia. The practice is counterproductive in this era of collaborative science.

Furthermore, review articles do not count towards impact metrics in China, so they do not help to advance an author's academic career. Information overload means that multidisciplinary, scholarly and timely reviews are more in demand than ever. Review authors need comprehensive knowledge, expert insight and outstanding inductive and deductive abilities. We therefore believe that this apparent discrimination is unjustified.

**Xiangyu Ma** *College of Preventive Medicine, Third Military Medical University, Chongqing, China.*  
**Zhiyuan Song** *Southwest Hospital, Third Military Medical University, Chongqing, China.*  
xymacq@hotmail.com

## US shutdown should spur other nations

What the recent US government shutdown has highlighted is just how much it contributes to the world's research infrastructure — with PubMed a striking example. By comparison, governments of other high-GDP countries contribute shamefully little.

The world urgently needs freely accessible, searchable, comprehensively referenced and cross-linked international repositories for the burgeoning amounts of untapped 'omics' data (lipidomics being one example). It is time for other wealthy countries to create and fund more such platforms.

**Julio Licinio, Ma-Li Wong** *South Australian Health and Medical Research Institute, and Flinders University, Adelaide, Australia.*  
julio.licinio@sahmri.com

## Riding shutdowns in developing world

The recent US government shutdown hit researchers in the Antarctic particularly hard (G. E. Hofmann *Nature* **502**, 431–432; 2013). But spare a thought for the feats of endurance shown by scientists from the developing world, who continue working in such harsh environments while contending with chronically unpredictable political factors in their own countries.

These scientists regularly battle long-term, government-induced setbacks to their research programmes. Regardless of their scientific importance, these studies do not receive widespread publicity when things go wrong — because such events are so frequent. But adverse environmental conditions offer only a limited logistical window for researchers, and natural systems will not wait for the resolution of political brawls.

Examples of such successful long-term research programmes

from developing-world scientists include Indian and Brazilian initiatives in the Antarctic (S. D. Gad *Curr. Sci.* **95**, 151; 2008, and P. Artaxo *et al. Tellus* **44B**, 318–334; 1992), and an ongoing subantarctic marine-mammal study by South Africans that has so far lasted for 40 years (M. N. Bester *et al. Afr. J. Mar. Sci.* **33**, 511–521; 2011).

**P. J. Nico de Bruyn** *University of Pretoria, South Africa.*  
pjndebruyn@zoology.up.ac.za

## Science is not about simple stories

Presenting science as a compelling story is becoming a popular way of communicating results — a technique that is guaranteed to capture the attention of the scientific community and the public. Although science needs great stories, stories are not science.

Storytelling glosses over uncertainties; methodological detail and even results can get lost in the narrative's overarching trajectory. By bounding over hurdles, it keeps the listener rooting for an amazing ending.

But we should not forget that the temptation to package results into a satisfying story has frequently led to the retraction of scientific papers.

**Jeroen Bergmann** *The Brain Sciences Foundation, Rhode Island, and Massachusetts Institute of Technology, USA.*  
jeroen.bergmann@media.mit.edu

### CORRECTION

The Correspondence 'Discard Soviet doctorate systems' (A. Gorobets *Nature* **503**, 39; 2013) wrongly implied that researchers in Russia, Ukraine, Belarus and other former Soviet states do not rely on peer-reviewed publications for career advancement. In fact, they do not rely on international peer-reviewed publications. Also, doctorate-degree candidates are not classified as 'students'.

## A cascade laser's random walk

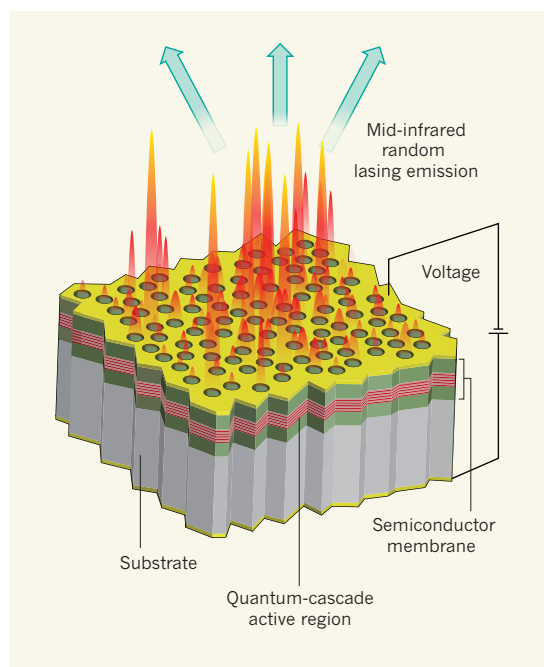
A random array of holes etched in a semiconductor structure, consisting of a periodic series of thin layers, has been demonstrated that emits mid-infrared laser radiation. The device could have sensing and imaging applications.

HUI CAO & STAFFORD W. SHEEHAN

Fifty-three years after their initial demonstration by Theodore Maiman<sup>1</sup>, lasers are now ubiquitous in our daily lives. We use them at the checkout counters of supermarkets, to read or write to DVDs, or just to surf the Internet (through fibre-optic communication systems). Lasers are also widely used in industrial settings and hospitals. Over the years, scientists and engineers have developed many types of laser using various materials for diverse applications. But despite such tremendous advances in laser technology, it is still a pleasant surprise to learn of developments that may pave the way for further cutting-edge applications. One example of this is a special class of laser — a random laser — operating in the mid-infrared that Liang *et al.* report<sup>2</sup> in *Advanced Materials*.

A laser has two basic components: a gain medium, which is a material that emits and amplifies light; and an optical cavity, a structure that traps light temporarily to make amplification efficient. The simplest laser optical cavity is made up of two mirrors facing each other, one highly reflective and the other partially transmitting, with a gain medium between them. Light is reflected back and forth between the mirrors, and each time it passes through the gain medium it is amplified, until it leaves through the partially transmitting mirror. A lasing threshold, beyond which laser light is emitted, is reached when the amplification of light caused by the gain medium compensates for the loss of light from the mirrors.

Unlike a conventional laser, a random laser uses a highly disordered structure to trap the light<sup>3,4</sup>. Although they might seem incompatible with lasers, disordered materials are common in our daily lives: milk, toothpaste, paper and clouds are just a few examples. They all consist of small particles or droplets that scatter light. When a light wave hits a particle, its direction of travel is changed randomly, and that particle is called a scattering centre. After propagating farther, the light wave will hit another scattering centre



**Figure 1 | An electrically pumped mid-infrared random laser.** Liang *et al.*<sup>2</sup> fabricated a random array of holes open to the air in a semiconductor membrane on a substrate. When an external electrical voltage is applied between two gold electrodes, electrons are injected from the top electrode to the quantum-cascade active region in the membrane, which subsequently emits and amplifies light. That light is scattered by the air holes, bouncing from one hole to another before escaping vertically or laterally. The multiple scattering and interference of light traps it temporarily inside the active region, enhancing light amplification. When light leakage and absorption are balanced by this amplification, lasing occurs.

and change direction again. In a system that consists of thousands of scattering centres, a light wave will be scattered many times and undergo a 'random walk'. Multiple scattering in this manner increases the path length of the light, in turn increasing its dwell time in a disordered structure. If a gain medium consisting of dye molecules is incorporated between scattering centres, light amplification is enhanced by multiple scattering, and lasing can occur.

Since the pioneering work of Nicolay Basov<sup>5</sup> and Vladilen Letokhov<sup>6</sup> in the 1960s, extensive studies<sup>3,4</sup> have shown that random lasing is a phenomenon that applies to numerous systems, including powders, colloids, polymers,

fibres and even biological tissues. Lasing frequencies range from the near-ultraviolet to the near-infrared. A major obstacle to the practical application of random lasers is the expensive source that is required to pump energy into the gain medium so that it can amplify the light. There are two pumping schemes: one is optical, involving absorption of light from another source (usually a laser at a different frequency); the other is electrical, such as injection of current from a battery. The second scheme is much cheaper and more convenient for many applications. However, most random lasers are optically pumped.

Liang and colleagues' work represents a major advance in random-laser technology. The authors pushed the lasing frequency to the mid-infrared, and achieved electrical pumping. To do this, they etched a random array of air holes into a semiconductor membrane (Fig. 1). Light propagating in the membrane is scattered from one hole to another, randomly changing direction at each one — rather like a drunk person wandering in a forest and changing direction whenever he encounters a tree, to use a common random-walk analogy. Because this structure is fabricated by lithography and chemical etching, the diameter, position and density of the holes can be accurately determined, allowing precise control of the scattering strength. When the scattering is sufficiently strong, random lasing occurs. The lasing threshold decreases with increasing air-hole density, because more holes help to trap light more effectively.

To push their lasing wavelength down to 10 micrometres, Liang *et al.* used a quantum cascade laser, which is an intrinsically electrically pumped device that provides strong amplification of light in the mid-infrared<sup>7</sup>. The membrane comprises a periodic series of thin layers of compound semiconductors. When an external voltage is applied, a representation of the energy levels of electrons in these layers would look like a staircase. Once an electron is at the top of the staircase, it begins to cascade down and emits a photon with every step that it descends. The quantum cascade laser's



emission frequency is determined by the energy drop at each step, which can be adjusted by changing the layer thickness.

Liang and co-workers' quantum cascade random laser raises hopes of new imaging applications. Many biochemical molecules have spectral 'fingerprints' in the mid-infrared range, a property that can be used for sensing and imaging applications — for example, in trace-gas sensing for pollution control and environmental monitoring; combustion diagnostics; and medical diagnostics (such as breath analysis). A crucial characteristic of random lasers is that the spatial coherence of their laser emission can be easily tuned<sup>8</sup>. Spatial coherence describes the correlation between waves at different points in space: sunlight shining through a cloud has low spatial coherence, whereas stars in the night sky have high spatial coherence. Low spatial coherence is desirable for parallel imaging, because light illuminating different spatial positions is uncorrelated, and so coherence artefacts such as crosstalk and speckle (random granular patterns) are avoided<sup>9</sup>.

To produce emission that has low spatial coherence, a random laser must have many modes (oscillating waves) lasing simultaneously<sup>8</sup>. Liang and colleagues' laser supports only a limited number of random-lasing modes, so its spatial coherence could still be high, although this has not been measured. A further increase in the laser's scattering strength would probably cause more modes to lase, thereby reducing the spatial coherence of the emission. However, because random-laser emission is non-directional, in this case a clever scheme would need to be developed for the efficient collection of total emission.

Looking to the future, this method of making mid-infrared random lasers could be extended to terahertz ( $10^{12}$  Hz) frequencies. In fact, lasers known as quasi-periodic, distributed-feedback quantum cascade lasers operating in the terahertz regime were constructed only three years ago<sup>10</sup>. Common packaging materials such as cardboard and plastics are transparent to terahertz radiation, making it useful for package inspection, quality control and non-destructive testing<sup>7</sup>. A combination of low spatial coherence and high radiance would also make a terahertz random laser suitable for applications such as high-speed parallel inspection. ■

**Hui Cao and Stafford W. Sheehan** are in the Department of Applied Physics, Yale University, New Haven, Connecticut 06520, USA.  
e-mail: hui.cao@yale.edu

5. Ambartsumyan, R. V., Basov, N. G., Kryukov, P. G. & Letokhov, V. S. *IEEE J. Quantum Electron.* **2**, 442–446 (1966).
6. Letokhov, V. S. *Sov. Phys. JETP* **26**, 1246–1251 (1968).
7. Gmachl, C., Capasso, F., Sivco, D. L. & Cho, A. Y.

- Rep. Prog. Phys.* **64**, 1533–1601 (2001).
8. Redding, B., Choma, M. A. & Cao, H. *Opt. Lett.* **36**, 3404–3406 (2011).
9. Redding, B., Choma, M. A. & Cao, H. *Nature Photon.* **6**, 355–359 (2012).
10. Mahler, L. *et al. Nature Photon.* **4**, 165–169 (2010).

## BIOCHEMISTRY

# Metal ghosts in the splicing machine

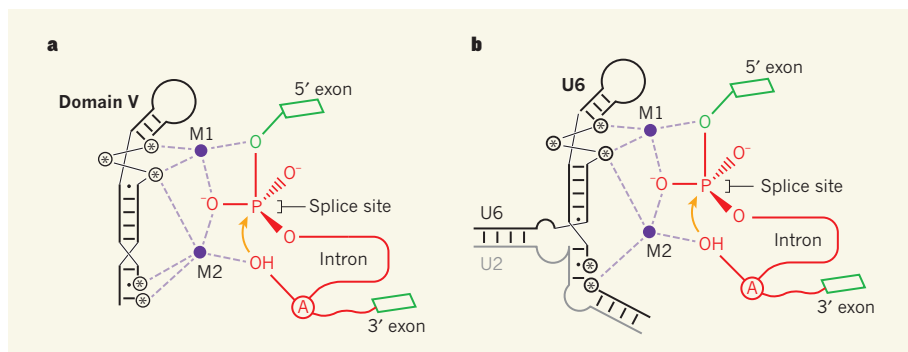
**Chemical analysis of the spliceosome's active site reveals that it is the RNA components of this enzyme complex that coordinate the catalytic metal ions responsible for production of a spliced messenger RNA. SEE ARTICLE P.229**

SCOTT A. STROBEL

**G**enes are transcribed as pre-messenger RNA molecules in which the coding exon segments are typically interrupted with non-coding sequences, or introns. To create a functional mRNA, the introns must be removed and the exons joined together in a process known as pre-mRNA splicing. The two phosphoryl-transfer reactions of splicing are catalysed by the spliceosome, a metallo-enzyme complex comprising five small nuclear RNAs (snRNAs) and dozens of core proteins<sup>1</sup>. The spliceosome uses magnesium ions at its active site as catalytic cofactors, and researchers have long been captivated by the question of whether the ligands for these metal ions are provided by the snRNAs or by the proteins, as this speaks to the evolutionary roots of the splicing process. On page 229

of this issue, Fica *et al.*<sup>2</sup> report that a specific constellation of phosphate oxygens in the U6 snRNA supplies the ligands that coordinate the catalytic metal cofactors in the active site. This finding establishes the concept that pre-mRNA splicing is fundamentally an RNA-catalysed reaction.

Fica and colleagues used an approach known as a metal-specificity switch to establish how the catalytic metal ions are bound within the spliceosome. They systematically introduced single atoms of sulphur in place of phosphate oxygens in the U6 snRNA of the spliceosome. Of the 20 positions tested, they found five substitutions that impaired splicing in the presence of magnesium ions ( $Mg^{2+}$ ), a metal that binds efficiently to oxygen but not to sulphur. The splicing activity of these impaired spliceosomes could be rescued by the addition of manganese ( $Mn^{2+}$ ) or cadmium ( $Cd^{2+}$ )



**Figure 1 | Active sites of group II introns and the spliceosome.** **a**, Group II introns are RNA elements that self-catalyse splicing — the removal of introns (red) and the joining of adjacent exon regions (green). They contain a hairpin structure called domain V; the phosphoryl-transfer reactions of splicing (the first of which is shown with an orange arrow) are catalysed by two divalent metal ions (M1 and M2) that are coordinated (dashed purple lines) to phosphate-group oxygen atoms of nucleotides (asterisks) in this region<sup>5</sup>, to a non-bridging oxygen at the splice site and to an OH group from an adenosine nucleotide (A) in the intron. **b**, Fica *et al.*<sup>2</sup> now show that RNA also catalyses these reactions during pre-mRNA splicing by the spliceosome — a metalloenzyme complex composed of small nuclear RNAs (snRNAs) and proteins. They show that the U6 snRNA of the spliceosome (which forms a hairpin structure similar to that of domain V) provides the phosphate oxygens that bind to two catalytic metal ions at the enzyme's active site, in a similar manner to the metal-ion coordination of group II introns. The adjoining U2 snRNA of the spliceosome is also shown.

1. Maiman, T. H. *Nature* **187**, 493–494 (1960).
2. Liang, H. K. *et al. Adv. Mater.* <http://dx.doi.org/10.1002/adma.201303122> (2013).
3. Cao, H. *Prog. Opt.* **45**, 317–370 (2003).
4. Wiersma, D. S. *Nature Phys.* **4**, 359–367 (2008).

ions — metals that bind more efficiently to sulphur than does magnesium. Such changes in metal specificity constitute a chemical signature that has been used to establish metal-ion binding to specific oxygens within other RNA active sites<sup>3</sup>.

Efforts to map metal-ion binding sites in the U6 snRNA have been made before using this approach, albeit on a less comprehensive scale (for example, see ref. 4). What sets Fica and colleagues' work apart is the triangulation of binding sites in U6 to the reactive groups in the pre-mRNA substrate. The authors establish that at least four of the five phosphate oxygens are coordinated to the catalytic Mg<sup>2+</sup> ions during the two phosphoryl-transfer reactions of splicing. They demonstrated this by preparing complexes with double and even triple sulphur-substituted spliceosomes and substrates. As the number of substitutions increased, they observed a requirement for greater thiophilicity (sulphur-binding preference) of the rescuing metal: Cd<sup>2+</sup> provided rescue but Mn<sup>2+</sup> did not.

Titration of the Cd<sup>2+</sup> concentration needed to achieve rescue suggests that there are two different catalytic metal ions within the active site (one coordinated to the nucleophilic hydroxyl and one coordinated to the leaving group), and that these metals interact with distinct phosphate oxygens in U6. Interestingly, no specificity switch was detected upon sulphur substitution in the RNaseH domain of Prp8, a protein at the core of the spliceosome that had been speculated to have a catalytic role<sup>5</sup>.

The inorganic architecture of the spliceosome active site that emerges from these studies is strikingly similar to that seen in the self-splicing group II introns<sup>6</sup> (Fig. 1). This subclass of introns are able to splice themselves out of the pre-mRNA without the need for the spliceosome, yet group II splicing and spliceosomal splicing involve the same two phosphoryl-transfer reactions and generate the same reaction intermediates and products. The crystal structure of a group II intron<sup>6</sup> revealed an active site with two catalytic metal ions coordinated to phosphate groups in the domain V helix of the RNA that are spatially and functionally equivalent to the phosphates that Fica *et al.* show are coordinated between the U6 snRNA and the pre-mRNA. The parallels between these molecules even extend to the stereochemistry of the oxygen atoms that serve as ligands for the two metal ions. Although it has long been known that the hairpin-loop structures of U6 and domain V are functionally similar, this work demonstrates the chemical equivalence between these two systems in their metal-ion coordination.

There are more than 100 proteins in the spliceosome, so what role do they have if RNA is the catalyst of pre-mRNA splicing? Some proteins are retained throughout the splicing process, whereas others associate

or dissociate at various steps in the reaction. Fica and colleagues' data do not exclude the possibility that some of the catalytic metal ligands are provided by protein functional groups. The authors tested one amino-acid candidate in spliceosomal factor Prp8 and obtained a negative result, but other segments of Prp8, particularly a loop close to the active site<sup>7</sup>, remain a possibility. In addition to the two catalytic divalent metals present in domain V of group II introns, this structure also contains two monovalent potassium (K<sup>+</sup>) ions, resulting in a cluster of four metal ions<sup>6</sup>. Does the inorganic architecture extend to the monovalent ions, or do proteins play the part of K<sup>+</sup> within the spliceosome? Whatever the answers to these questions, it seems that the main role of the proteins is not catalytic, but that they are instead involved in regulation, in determining splice-site specificity and as a scaffold to promote the proper conformation of the snRNAs.

Although the group II intron structure is invaluable, we still lack high-resolution crystal structures of the spliceosome in any of its various reaction states. Such a structure would reveal the full relationship of U6 and its associated proteins to the catalytic metals and pre-mRNA substrate. The complexity, heterogeneity and size of the spliceosome have so far precluded structural analysis at high enough resolution to extract biochemical mechanistic information, but cryo-electron microscopy images continue to improve<sup>8</sup>. There is also a need to confirm the mechanistic parallels between the spliceosome and the group II self-splicing system. The biochemical data available on U6 have outpaced those for domain V, so it has not been tested whether the metals observed in domain V show a metal-specificity signature consistent with a direct role in

catalysis, as has now been demonstrated for the metal ions in U6.

In the early days of our understanding of self-splicing RNAs, chemist Thomas Cech posed the question: is RNA-based catalysis a general phenomenon or a rare molecular fossil<sup>9</sup>? For many years, the study of RNA catalysis was confined to introns in genomes outside the nucleus and in non-model organisms, including yeast mitochondria and ciliated protozoa. But the crystal structure of the ribosome<sup>10</sup> revealed that RNA provides the active site responsible for protein synthesis across all forms of life. The functional demonstration by Fica *et al.* that pre-mRNA splicing is catalysed within an RNA metalloenzyme active site clearly shows that RNA-based catalysis is a fundamental and ubiquitous process involved at several steps in the conversion of genetic information to functional proteins. ■

Scott A. Strobel is in the Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520, USA.

e-mail: scott.strobel@yale.edu

1. Sontheimer, E. J., Sun, S. & Piccirilli, J. A. *Nature* **388**, 801–805 (1997).
2. Fica, S. M. *et al.* *Nature* **503**, 229–234 (2013).
3. Fredericksen, J. K. & Piccirilli, J. A. *Methods* **49**, 148–166 (2009).
4. Yean, S. L. *et al.* *Nature* **408**, 881–884 (2000).
5. Pena, V., Rozov, A., Fabrizio, P., Lüthmann, R. & Wahl, M. C. *EMBO J.* **27**, 2929–2940 (2008).
6. Marcia, M. & Pyle, A. M. *Cell* **151**, 497–507 (2012).
7. Galej, W. P., Oubridge, C., Newman, A. J. & Nagai, K. *Nature* **493**, 638–643 (2013).
8. Golas, M. M. *et al.* *Mol. Cell* **40**, 927–938 (2010).
9. Cech, T. R. *Cell* **44**, 207–210 (1986).
10. Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. *Science* **289**, 920–930 (2000).

This article was published online on 6 November 2013.

## SOLAR SYSTEM

# Russian skyfall

The recent entry of a 20-metre-wide celestial rock into Earth's atmosphere offered both a spectacular show and a source of invaluable data that advance our understanding of high-velocity impacts. **SEE LETTERS P.235 & P.238**

NATALIA ARTEMIEVA

Two papers in this issue focus on the scientific reconstruction of the asteroid-impact event that was observed on 15 February 2013 in Chelyabinsk, Russia — the largest impact event on Earth since the Tunguska blast of 1908. Borovička *et al.*<sup>1</sup> (page 235) extract substantial scientific information from several low-quality video recordings. Meanwhile, by analysing infrasound airwaves and the brightness of the light

flash generated by the impact, as well as the damage caused to Earth's surface, Brown *et al.*<sup>2</sup> (page 238) estimate the total energy of the event to have been 400–600 kilotons of trinitrotoluene (1 kt of TNT is equivalent to  $4.185 \times 10^{12}$  joules of energy), and say that such events occur more often than previously thought.

There are three fundamental physical processes involved as meteoroids — asteroids' smaller counterparts — enter the atmosphere. First, atmospheric drag decelerates meteoroids from their initial velocity of about





Figure 1 | The chunk of the Chelyabinsk asteroid that was pulled from Lake Chebarkul.

20 kilometres per second to a free-fall velocity ranging from several centimetres to hundreds of metres per second, depending on the meteoroid's size. Second, atmospheric shock waves heat the air surrounding the meteoroid and make the body visible as a 'bolide'. The waves' thermal radiation evaporates the meteoroid, which may lose more than 90% of its initial mass before it hits the ground. Third, extremely high dynamic pressure breaks the meteoroid into smaller fragments, which in turn are subjected to even higher degrees of deceleration and evaporation.

In his letter to this journal<sup>3</sup> in 1908, W. F. Denning described the night of 30 June in Bristol: "The firmament was abnormally luminous, with a very strong glow all over the north at midnight." Two other letters on the same page (and others in *Nature's* next issue) reported bright nights in Berlin, London and Prague. On the same day, atmospheric disturbances and seismic signals were registered across Eurasia, with the strongest reported in Russian Siberia. In 1927, a huge area of devastated forest was discovered near the Podkamennaya Tunguska River, also in Siberia. It took decades to connect all of these observations with a cosmic-body impact, now known as the Tunguska event, which carried a total energy of 10–15 Mt TNT (ref. 4). Taken together with the hypothesis<sup>5</sup> that an asteroid impact caused the mass extinction at the Cretaceous–Palaeogene geological boundary 66 million years ago, the Tunguska event highlighted a new threat to the human race and a challenge for astronomers: to predict asteroid impacts in advance.

So far, some 90% of near-Earth objects, such as comets and asteroids, larger than 1 kilometre in diameter have been monitored<sup>6</sup>. Smaller, Tunguska-like bodies are much more difficult to detect. However, the 2012 DA<sub>14</sub> near-Earth asteroid, with a diameter of about 30 metres, was discovered in February 2012.

A year later, in Tucson, Arizona, I was watching NASA's news on television to find out how to observe the asteroid during its expected closest approach of some 27,700 km the next day. Several minutes later, breaking news came from a radio tuned to a Russian station, announcing that a superbolide had exploded over Chelyabinsk.

My first thought was that the Russian media had misinterpreted NASA's news. But it was not a mistake. Thousands of people observed a bright flash in the sky followed by a powerful sonic boom that destroyed windows across an area of about 5,000 km<sup>2</sup>. More than 1,500 people were injured, mainly by broken glass. Amateur video recordings of the event allowed Borovička and colleagues to reconstruct the body's trajectory and its fragmentation history: the entry angle was unusually low, approximately 17°; the observed trajectory in the atmosphere exceeded 250 km in length; several light flashes occurred below an altitude of 40 km; and hundreds of fragments were dispersed along the trajectory.

Borovička *et al.* say that the largest fragment (which they name F1) would have touched down in Lake Chebarkul. The recent recovery of this fragment (Fig. 1), which weighs some 600 kg, is the best prize for the scientists. The authors also reconstruct the Chelyabinsk asteroid's space biography and connect its origin with the 86039 near-Earth asteroid, rather than the long-anticipated 2012 DA<sub>14</sub> asteroid, which passed by Earth almost unnoticed, thanks to its loud precursor. Brown and colleagues estimate the pre-atmospheric diameter of the asteroid to be in the range of 17–20 m, with a total mass of about 10<sup>7</sup> kg. Only a tiny fraction of this mass, less than 0.01%, has been recovered on Earth's surface.

Both teams used standard models of impacts to make their calculations. Although these models have been validated for numerous small meteoroids, they may fail for meteoroids

that are larger than 10 m in diameter. For such objects, the collective behaviour of fragments plays a crucial part: evaporation could be more intense and account for the final mass deficiency, whereas dynamic pressure could be much lower, explaining why large fragments survive. It is to be hoped that the Chelyabinsk event will prompt researchers to refine their models.

How common are Tunguska- and Chelyabinsk-like events — should we expect other guests from outer space any time soon? Standard methods<sup>7</sup> predict a Tunguska-like event once every millennium, and a Chelyabinsk-like event once a century. Brown and colleagues are more pessimistic, suggesting a frequency that is tenfold higher for both types of event. We will probably have to wait some 100 years to find out.

And are astronomers capable of predicting the impacts of meteoroids a few metres in diameter? Probably not: such objects are too small to be detected in space by modern telescopes. One successful prediction was that of asteroid 2008 TC<sub>3</sub>, which was discovered by accident a day before it struck Earth's atmosphere, and recovered later in Sudan as the Almahata Sitta meteorite<sup>8</sup>. If the trajectory, size and composition of a body are known, physicists can predict its behaviour — whether it will be harmlessly vaporized and fragmented in the atmosphere or whether it will cause serious damage to the planet's surface (such as shock waves, wildfires and impact craters).

Finally, can humans minimize the risk of impact-related injuries? Yes they can, by learning a few basic facts about meteoroids and their interaction with the atmosphere — just as people living in seismically active areas are prepared for earthquakes. A meteoroid impact could happen at any time, anywhere on Earth. If you see a bright flash in the sky, do not panic; stay away from the windows; and find a secure spot to hide. If possible, use your mobile phone to take pictures and videos, and time the interval between the flash and the sonic boom (it may take up to a few minutes). Without your observations, a future impact event might remain as enigmatic as that at Tunguska. ■

**Natalia Artemieva** is in the Planetary Science Institute, Tucson, Arizona 85719, USA, and at the Institute for Geosphere Dynamics, Russian Academy of Sciences, Moscow. e-mail: artemieva@psi.edu

1. Borovička, J. *et al.* *Nature* **503**, 235–237 (2013).
2. Brown, P. G. *et al.* *Nature* **503**, 238–241 (2013).
3. Denning, W. F. *Nature* **78**, 221 (1908).
4. Shuvalov, V. V. & Trubetskaya, I. A. *Solar Sys. Res.* **41**, 220–230 (2007).
5. Alvarez, L. W., Alvarez, W., Asaro, F. & Michel, H. V. *Science* **208**, 1095–1108 (1980).
6. <http://neo.jpl.nasa.gov/stats>
7. Michel, P. & Morbidelli, A. *Meteor. Planet. Sci.* **42**, 1861–1869 (2007).
8. Jenniskens, P. *et al.* *Nature* **458**, 485–488 (2009).

## STEM CELLS

# Differentiated cells in a back-up role

Two independent studies show that, if push comes to shove, differentiated cells of the stomach and lung can act as adult stem cells, generating various cell types of the tissues, including a pool of stem cells. [SEE ARTICLE P.218](#)

TUSHAR J. DESAI & MARK A. KRASNOW

When functional cells die, they are soon replaced. In most cases, the replacement cells arise either from the division of surviving mature cells of the same class or from the division and differentiation of tissue stem cells. But what happens when resident stem cells are selectively depleted? Two papers, one by Tata *et al.*<sup>1</sup> on page 218 of this issue and the other by Stange *et al.*<sup>2</sup> published in *Cell*, find that following depletion of stem cells in the stomach or lung, stem-cell function can be recovered through a surprising back-up function provided by specific differentiated cells in each tissue.

The stomach and lung are lined by a single layer of cells of various types that are continually replaced throughout life as they become damaged or die. In the central part of the stomach, the cells lining periodic evaginations called crypts are found in a specific distribution, organized by class (Fig. 1a). It is thought that cells in individual crypts are maintained by rapidly dividing tissue stem cells that reside just above the crypt's midpoint and whose daughter cells spread in both directions,

differentiating into cells of various classes<sup>3</sup>.

Stange and colleagues, working in mice, find that the base of the crypts contains cells that express Troy, a marker of intestinal stem cells. Using genetic techniques to 'pulse-label' these cells in a permanent and heritable manner at a low frequency, they occasionally find crypts in which all cells are derived from a Troy-expressing cell whose progeny slowly spread up from the base. When the authors destroyed the tissue stem cells, however, Troy-expressing cells executed this stem-cell function much more rapidly and in many more crypts. Remarkably, the Troy-expressing cells are a type of fully mature secretory cell called a chief cell, which maintains its differentiated identity even while performing its stem-cell function. Because their regenerative function is activated following depletion of the tissue stem-cell population, chief cells can be considered reserve stem cells.

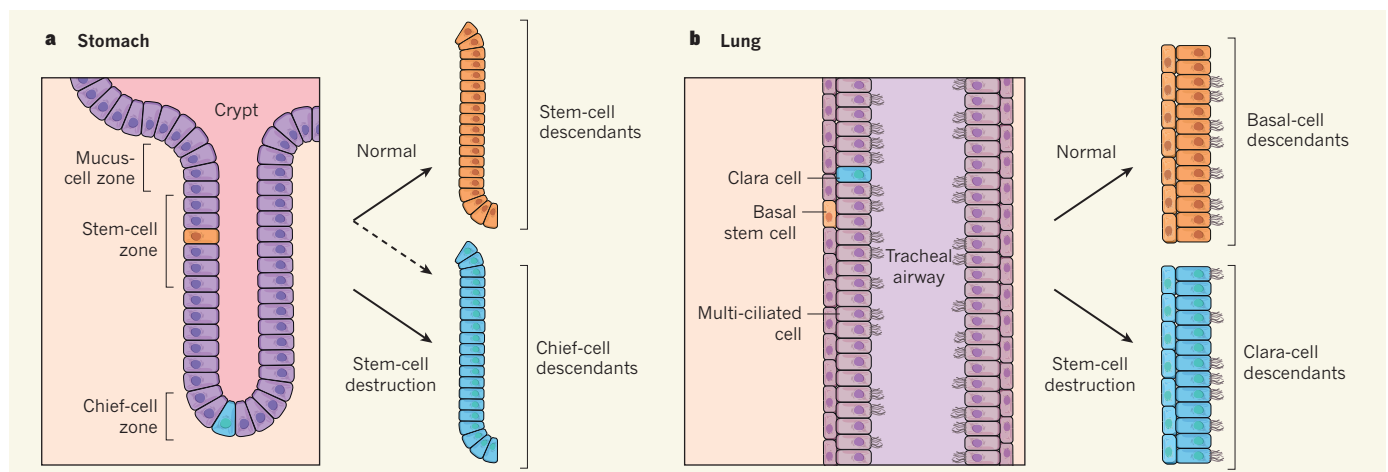
Tata *et al.* independently demonstrate that differentiated airway secretory cells known as Clara cells can contribute to regeneration in the lung. Previous work showed<sup>4,5</sup> that undifferentiated basal cells in the mouse trachea replenish the stock of secretory and multi-ciliated cells, which produce and clear airway

mucus, respectively. In the present paper<sup>1</sup>, the investigators pulse-labelled mature secretory cells en masse before specifically killing basal cells. Surprisingly, they later found the lineage mark they had introduced before basal-cell destruction in newly arising basal cells. But Tata and co-workers' bulk-labelling strategy is a potential caveat, because it may have inadvertently marked some original basal cells that escaped destruction. It would be valuable to conduct studies using a sparse-labelling strategy, to trace the behaviour of individual secretory cells.

These authors go on to show that the marked basal cells, presumably descendants of labelled mature secretory cells, function as stem cells, renewing both multi-ciliated and secretory cell types (Fig. 1b). Because their progenitor activity is elicited only after elimination of basal stem cells, tracheal Clara cells can also be considered reserve stem cells.

Although the differentiated Clara cells of the lung and chief cells of the stomach each give rise to multiple cell types, the routes they take are different. Clara cells generate replacement stem cells, whereas chief cells apparently bypass this requirement and are themselves stem cells. However, lower down in the airway tract, Clara cells seem to be stem cells, renewing themselves and multi-ciliated cells without first becoming basal cells<sup>6</sup>. Conversely, chief cells also seem to generate stem cells, albeit indirectly, because their descendants eventually replace all crypt cells, including the resident stem-cell populations. Thus, despite taking different routes, these mature cells share the potential to generate both differentiated cells and stem cells.

The two papers challenge the primacy of undifferentiated, resident stem cells, given that mature cells can substitute for their function and even make new ones. In other tissues, differentiated cells may similarly provide a



**Figure 1 | When mature cells function as reserve stem cells.** **a**, In the crypts of the stomach, undifferentiated adult stem cells (such as the orange cell) occupy the stem-cell zone, whereas functional, enzyme-secreting chief cells (blue) reside at the base. During cellular turnover, stem cells generate all cell types of the crypt (orange). But Stange *et al.*<sup>2</sup> find that if stem cells are destroyed, mature chief cells assume a stem-cell role, renewing even the depleted stem cells. Chief cells infrequently renew crypts in the absence of obvious injury (dotted arrow). **b**, The adult tracheal stem cells called basal cells (orange) replenish the complement of secretory Clara cells and multi-ciliated cells under normal physiological conditions and after injury. When Tata *et al.*<sup>1</sup> destroyed these stem cells, Clara cells became activated to regenerate basal cells, which resumed the task of maintaining tracheal cell types.



reserve stem-cell function when the primary renewal mechanisms are inadequate<sup>7,8</sup>. The new studies also raise questions, such as what reprogramming factors regulate stem-cell behaviour in mature cells, and whether reversion to an undifferentiated state is an obligate step. Also, which cells generate the primary stem-cell population in a tissue? And how is an appropriate balance between mature cells and different types of stem cells within a tissue maintained? Tata *et al.* provide evidence that, in the trachea, contact between Clara cells and basal cells or short-range inhibition of Clara-cell dedifferentiation by basal cells may play a part.

The pursuit of these questions may have implications for regenerative medicine, given

that there is an intrinsic appeal to the shorter path in redirecting differentiation of a mature cell instead of starting from scratch with an undifferentiated stem cell. Equally important is the possibility that these 'reserve' programs can be activated in differentiated cells *in vivo* by extrinsic signals. This would eliminate the need to introduce cellular reprogramming factors, and thereby avoid the attendant risk of promoting cancer through this form of potential therapy. ■

**Tushar J. Desai** is in the Department of Medicine, Division of Pulmonary and Critical Care, and **Mark A. Krasnow** is in the Department of Biochemistry and Howard Hughes Medical Institute, Stanford University

School of Medicine, Stanford, California 94305-5307, USA.

e-mails: [tdesai@stanford.edu](mailto:tdesai@stanford.edu); [krasnow@stanford.edu](mailto:krasnow@stanford.edu)

1. Tata, P. R. *et al.* *Nature* **503**, 218–223 (2013).
2. Stange, D. E. *et al.* *Cell* **155**, 357–368 (2013).
3. Karam, S. M. & Leblond, C. P. *Anat. Rec.* **236**, 259–279 (1993).
4. Hong, K. U., Reynolds, S. D., Watkins, S., Fuchs, E. & Stripp, B. R. *Lung Cell. Mol. Physiol.* **286**, L643–L649 (2004).
5. Rock, J. R. *et al.* *Proc. Natl Acad. Sci. USA* **106**, 12771–12775 (2009).
6. Rawlins, E. L. *et al.* *Cell Stem Cell* **4**, 525–534 (2009).
7. Yanger, K. & Stanger, B. Z. *Dev. Dyn.* **240**, 521–529 (2011).
8. Xu, X. *et al.* *Cell* **132**, 197–207 (2008).

This article was published online on 6 November 2013.

## QUANTUM PHYSICS

# The right ambience for a single spin

**Long-lived single electron spins are crucial for quantum computation and for understanding spin dynamics. A remarkably long lifetime — of the order of minutes — has now been obtained for a solid-state system. SEE LETTER P.242**

MICHAEL E. FLATTÉ

Anyone who has had a pleasant dinner at a favourite restaurant ruined by noisy neighbours understands the disruption caused by too much interaction with one's environment. Most electronic spins in a solid are also buffeted by myriad naturally occurring 'noises', including nearby fluctuating electronic motion (spin–orbit interactions), interactions with other electronic or nuclear spins, and the mechanical motion of ions. The resilience of single-spin dynamics to these environmental effects is quantified by the spin coherence time or the closely related zero-field spin lifetime. Just as one might make modifications to soundproof a restaurant to improve the ambience for diners, so reducing the noises influencing a single spin by lowering the temperature, eliminating nuclear spins, and choosing solids made up of light atoms that have weak spin–orbit interactions, leads to long spin coherence times. Unfortunately, these methods also limit the materials in which long spin coherence times can be observed. But a deaf diner is impervious to noisy neighbours, and on page 242 of this issue, Miyamachi *et al.*<sup>1</sup> demonstrate an approach to making a single spin deaf to the dominant noises around it.

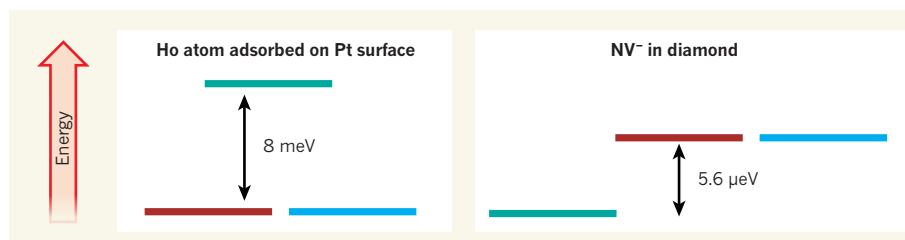
The system studied by Miyamachi and colleagues is a single holmium (Ho) atom adsorbed on the surface of platinum (Pt). The

Ho atom has an electronic spin of 8, and its lowest-energy spin states correspond to the spin pointing towards the surface or away from the surface; these two spin states are degenerate (of equal energy). A dominant source of noise for the electronic spin of an atom adsorbed on the surface of a metal comes from a passing conduction electronic spin, which interacts with the adsorbed spin and changes its orientation, transferring one quantum of angular momentum. When this occurs, the spin orientation of the adsorbed spin also changes, so the rate of this process can limit the coherence time of the adsorbed spin.

Lengthening the spin lifetime by reducing the interaction with the environment has

been demonstrated for single spins on metals, by building an insulating barrier between the adsorbed spin and the metal underneath<sup>2</sup>. In Miyamachi and colleagues' experiment, the Ho atom is adsorbed directly on a Pt surface (technically known as the (111) surface) chosen so that all the Pt surface atoms are arranged in regular, repeating equilateral triangles. The Ho atom sits in the centre of one of those triangles, and from its vantage point the surface would look the same if the entire surface were rotated 120° around it. Ordinarily, the presence of these neighbouring Pt atoms, combined with the spin–orbit interaction, would push the adsorbed spin into a quantum-mechanical state that is a superposition of the two low-energy (up and down) states, corresponding to a non-degenerate ground state for the Ho atom's electronic spin that has a vanishingly small spin orientation, and thus a short spin lifetime.

Here, however, it is this three-fold symmetry (so called because three rotations of 120° bring the surface back to its original configuration) that deafens the spin to its surroundings. The authors showed that for this geometric position of the Ho atom, and for the Ho atom's spin of 8, a transition from the Ho atom spin pointing away from the surface to it pointing towards the surface is not caused by the



**Figure 1 | Energy-level structure of spin systems.** The energy splitting between the two degenerate (equal energy) spin states and a third spin state for a holmium (Ho) atom on a platinum (Pt) surface is three orders of magnitude larger than for a nitrogen vacancy (NV<sup>-</sup>) centre in diamond. Furthermore, the two degenerate states are the ground states of the system, whereas the ground state of the NV<sup>-</sup> centre is a single state. This energy-level structure and large energy splitting for Ho on Pt was shown by Miyamachi *et al.*<sup>1</sup>, for temperatures corresponding to energies much less than the splitting, to eliminate spin processes that would reduce the lifetime of the electronic spin of the Ho atom.

presence of the neighbouring atoms, and also cannot directly occur through an interaction with a passing electron's spin — the Ho atom spin is oblivious to those interactions.

The transition remains possible with two spin-flips with passing electrons, but this requires the Ho spin to be in an intermediate orientation between the two spin-flips, and being in such an orientation costs a lot of energy. For Ho on Pt(111), the most important intermediate orientation is about 8 millielectronvolts (or, equivalently, 100 kelvin) above the ground state (Fig. 1). Thus, if the temperature is much less than 100 K, the two-spin-flip transition is extremely unlikely. As a result, at a temperature of 1.1 K, Miyamachi and colleagues measured a spin lifetime that exceeded 6 minutes — a remarkably long value for any solid-state spin system.

How does the lifetime of the Ho electronic spin compare with that of single spins in semiconductors? Can a similar energy-level structure be obtained, with the spin-flip of a passing electron unable to cause a transition between the two degenerate lowest-energy states? This configuration occurs for a manganese atom doped into gallium arsenide. In the bulk of gallium arsenide, a manganese atom has a three-fold (angular momentum 1) degenerate ground state, but the application of an electric field<sup>3</sup> or a strain field<sup>4</sup> can push one state much higher in energy, producing a spin-state level structure that should have long coherence times. Spin coherence times for manganese in gallium arsenide, however, do not exceed 10 nanoseconds<sup>5</sup> because of interactions with nuclear spins.

Another interesting comparison is with an electronic spin system called the nitrogen vacancy (NV<sup>-</sup>) centre in diamond<sup>6</sup>, in which a nitrogen atom and a vacancy replace two neighbouring carbons. That spin centre also has three-fold symmetry, and has two degenerate states. Therefore, the fundamental energy-level structure looks similar to that of Ho on Pt(111). The degenerate states of the NV<sup>-</sup> centre, however, are not ground states; the ground state is a single state and is split in energy from the other two states by only 5.6 microelectronvolts (Fig. 1). Thus, to keep long spin coherence times (a few milliseconds at room temperature<sup>7,8</sup>), the spin-orbit interaction must be very small.

Miyamachi and colleagues' exceptionally long lifetimes for single spins adsorbed on a metal strongly support the view that new single-spin candidates with improved fundamental properties can still be found, by careful consideration of the geometry and symmetry of the single spin within its environment. For a long-lived single spin, the effective coherence time can also be dramatically enhanced by careful selection and application of pulses of radiation<sup>9</sup>. To take full advantage of the long spin lifetimes of Ho on Pt(111) for spin-based computation or

temporary information storage, high-speed manipulation techniques must be developed for controlling the single-spin dynamics, such as have been demonstrated for single spins in diamond<sup>10</sup>. ■

**Michael E. Flatté** is in the Optical Science and Technology Center and the Department of Physics and Astronomy, University of Iowa, Iowa City, Iowa 52242, USA.  
e-mail: michael\_flatte@mailaps.org

1. Miyamachi, T. *et al.* *Nature* **503**, 242–246 (2013).
2. Heinrich, A. J., Gupta, J. A., Lutz, C. P. & Eigler, D. M. *Science* **306**, 466–469 (2004).

3. Tang, J.-M., Levy, J. & Flatté, M. E. *Phys. Rev. Lett.* **97**, 106803 (2006).
4. Yakunin, A. M. *et al.* *Nature Mater.* **6**, 512–515 (2007).
5. Myers, R. C. *et al.* *Nature Mater.* **7**, 203–208 (2008).
6. Jelezko, F., Gaebel, T., Popa, I., Gruber, A. & Wrachtrup, J. *Phys. Rev. Lett.* **92**, 076401 (2004).
7. Bar-Gill, N., Pham, L. M., Jarmola, A., Budker, D. & Walsworth, R. L. *Nature Commun.* **4**, 1743 (2013).
8. Balasubramanian, G. *et al.* *Nature Mater.* **8**, 383–387 (2009).
9. Dobrovitski, V. V., Fuchs, G. D., Falk, A. L., Santori, C. & Awschalom, D. D. *Annu. Rev. Condens. Matter Phys.* **4**, 23–50 (2013).
10. Fuchs, G. D., Dobrovitski, V. V., Toyli, D. M., Heremans, F. J. & Awschalom, D. D. *Science* **326**, 1520–1522 (2009).

## BIODIVERSITY

# The ecological deficit

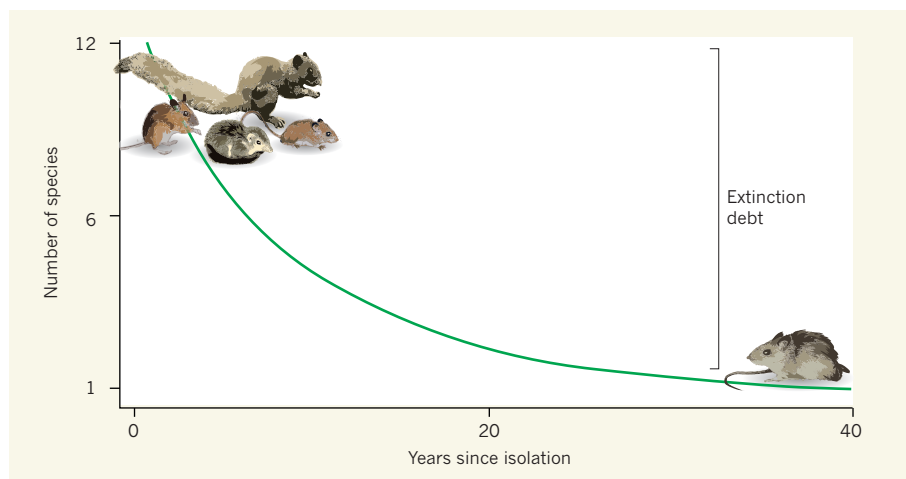
**The almost complete extinction of small mammals in forest islands within 25 years of the construction of a reservoir that fragmented the habitat provides a striking example of delayed biodiversity loss.**

ANDREW GONZALEZ

Global rates of extinction are set to peak this century<sup>1</sup>. But it is often difficult to predict when and where species losses will occur, and this hampers efforts to prevent them. Writing in *Science*, Gibson *et al.*<sup>2</sup> provide an example of things to come, identifying a combination of habitat fragmentation and the arrival of an invasive species as the cause of the extinction of an entire community of small mammals. The extinction sequence was revealed by repeated surveys of the mammalian fauna inhabiting a set of tropical forest islands in southern Thailand that were isolated in 1986 following inundation of the forest to

create the Chiew Larn reservoir. A mathematical model of the biogeography of these islands allowed the authors to estimate the extinction half-life of the mammal communities to be around 14 years — a striking example of an extinction debt.

Extinction debt is the idea that species may continue to be lost long after an initial disturbance<sup>3,4</sup>. Decades of research suggest that small islands of habitat cannot sustain the rich diversity of species they harboured at the time of isolation. Instead, diversity on habitat islands decays exponentially over time in a process called community relaxation<sup>5</sup>. The number of species lost from the start to the end of the relaxation process is the



**Figure 1 | Isolation leads to extinction.** Gibson *et al.*<sup>2</sup> document that within 25 years of the creation of islands by the construction of the Chiew Larn reservoir in Thailand, the diversity of native small mammals on the islands had decreased from up to 12 species to just one. On most of the islands, the only remaining small mammal was the invasive species *Rattus tiomanicus*, which the authors suggest contributed to the extinctions of native species.



size of the extinction debt (Fig. 1).

According to theory<sup>5</sup>, the rate at which an extinction debt is paid depends on the balance between the rate of extinction and the rate of immigration of species to habitat islands. In a newly isolated forest island, the rate of immigration may be so low that it cannot compensate for the high rate of local extinction.

Gibson and colleagues compared the number of mammalian species on 16 islands in the Chiew Larn reservoir with the biodiversity at sites on the nearby mainland, which contained up to 12 species of small mammal. They found that, after only five years of isolation, the smallest islands (those less than 10 hectares in area) harboured one to three native species, and that after 25 years even the largest islands (up to 56 ha) had fewer than five species. Surprisingly, they found that the larger the habitat island, the faster the loss of species.

The authors show that the effects of forest fragmentation were exacerbated by an invasive species. They identify the Malayan field rat (*Rattus tiomanicus*) as the culprit because on many of the islands it was the only small mammal remaining. This species is not naturally found in undisturbed tropical forests in this region, but it is common in human-transformed environments, such as agricultural fields and villages. When *R. tiomanicus* is common it often has devastating effects on native fauna. The authors believe that it may have spread to the islands after the area was inundated, and so was a major factor that accentuated the extinction-immigration imbalance and contributed to the eventual collapse of the island communities.

This study is exceptional because it combines field observations with an insightful mathematical model to reveal the process of biodiversity loss. It builds on long-term controlled experiments on several continents that revealed that habitat fragmentation causes extinction debts<sup>6</sup>. We now have accumulated evidence from field surveys and experiments that fragmentation has a lasting legacy. But further research is needed to examine how different drivers of extinction — such as climate change and exotic species — interact with habitat fragmentation to affect the rate and extent of an extinction debt. If we are to make accurate predictions of biodiversity loss, we must update extinction theory to account for the many effects humans have on biodiversity<sup>7</sup>.

An important but often overlooked consequence of an extinction debt is the impact it can have on ecosystem functions and services. This parallel idea of an ecosystem-function debt<sup>8</sup> arises from research showing that biodiversity influences the way ecosystems function. Consider, for example, the delayed effects of forest fragmentation on the diversity of wild pollinators and the economic legacy of a declining pollination service on agriculture and wild plant populations<sup>9</sup>. Or consider as well that habitat fragmentation may ultimately

result in the emergence of novel diseases carried by animals that can successfully invade a landscape only after substantial extinction has taken place<sup>10</sup>. These ecosystem impacts connect the extinction debt to human well-being.

So what can be done to minimize the causes and consequences of extinction debts? In a landscape composed of many habitat fragments, the creation of a reserve network can help to rebalance local extinction with immigration. Network science suggests that the size and connectivity of reserve networks can be designed to optimally maintain biodiversity<sup>11</sup> and sustain ecosystem services in the long term, while taking into account real-world economic constraints. In some cases, habitat connectivity can be restored by providing corridors to connect habitats; in others, creating close clusters of habitat fragments can achieve the same goal. Engineering the configuration of landscapes to minimize the effects of extinction debts is not the stuff of science fiction, but much more research is needed before we will know how best to do it.

Understanding extinction debts will improve our ability to predict when biodiversity loss will peak this century. Gibson and colleagues have provided a powerful and poignant

example of the science underlying this process, and their findings serve as a reminder that many extinctions are still to come. ■

**Andrew Gonzalez** is at the Quebec Centre for Biodiversity Science, and in the Department of Biology, McGill University, Montreal, Quebec H3A 1B1, Canada.

e-mail: [andrew.gonzalez@mcgill.ca](mailto:andrew.gonzalez@mcgill.ca)

1. Pimm, S. L. & Raven, P. *Nature* **403**, 843–845 (2000).
2. Gibson, L. *et al. Science* **341**, 1508–1510 (2013).
3. Tilman, D., May, R. M., Lehman, C. L. & Nowak, M. A. *Nature* **371**, 65–66 (1994).
4. Mouquet, N., Matthiessen, B., Miller, T. & Gonzalez, A. *PLoS ONE* **6**, e17567 (2011).
5. Diamond, J. M. *Proc. Natl Acad. Sci. USA* **69**, 3199–3203 (1972).
6. Tollefson, J. *Nature* **496**, 286–289 (2013).
7. Hylander, K. & Ehrlén, J. *Trends Ecol. Evol.* **28**, 341–346 (2013).
8. Gonzalez, A. & Charnet, E. J. *J. Anim. Ecol.* **71**, 594–602 (2002).
9. Tschirntke, T., Klein, A.-M., Kruess, A., Steffan-Dewenter, I. & Thies, C. *Ecol. Lett.* **8**, 857–874 (2005).
10. Li, S., Hartemink, N., Speybroeck, N. & Vanwambeke, S. O. *et al. PLoS ONE* **7**, e39612 (2012).
11. Cabeza, M. & Moilanen, A. *Trends Ecol. Evol.* **16**, 242–248 (2001).

## HIV

# Antibodies advance the search for a cure

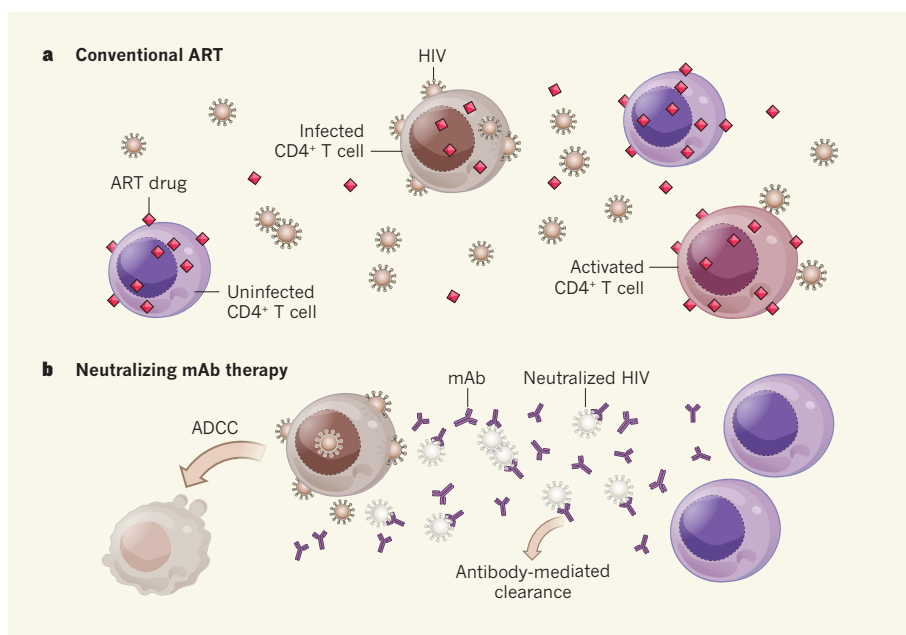
**Efforts to make a prophylactic HIV vaccine have identified monoclonal antibodies that potentially suppress viral replication. Studies in monkeys show that these reagents effectively treat HIV infection. [SEE ARTICLE P.224 & LETTER P.277](#)**

LOUIS J. PICKER & STEVEN G. DEEKS

A major breakthrough from the HIV-vaccine research community in recent years was the isolation and characterization of novel antibodies from HIV-infected people that have the remarkable ability to efficiently neutralize most circulating HIV strains<sup>1,2</sup>. The antibodies' mechanism of action involves the recognition and blockade of evolutionarily conserved, functionally crucial structures of the HIV viral envelope. These unusual antibodies have reinvigorated the effort to develop an antibody-based prophylactic HIV vaccine by defining effective human antibody responses to the virus and providing a 'map' for reverse engineering of vaccines that recapitulate these responses. Because these antibodies develop infrequently, emerge only after many years of HIV infection and are characterized by a high degree of mutation<sup>1,2</sup>, this task is not likely to be quickly accomplished. But that

does not necessarily mean that the clinical benefit of these antibodies is relegated to the distant future. On the contrary, two reports in this issue, by Barouch *et al.*<sup>3</sup> (page 224) and Shingai *et al.*<sup>4</sup> (page 277), demonstrate that combinations of such antibodies drastically reduce virus levels in chronically infected rhesus macaques, bolstering the hope that such therapies might be effective in humans.

The two groups of investigators examined the ability of several of these potent and broadly HIV-neutralizing human monoclonal antibodies (mAbs) to suppress viraemia — the presence of the virus in the bloodstream — in rhesus macaques chronically infected with pathogenic (AIDS-causing) simian-human immunodeficiency virus (SHIV) hybrids that use the HIV envelope for cell entry. Their findings were remarkably consistent: administration of cocktails of two or more mAbs (directed at two different structural regions, the CD4-binding site and V3-loop/N332



**Figure 1 | Differences between conventional ART and mAb therapy.** **a**, Conventional antiretroviral therapy (ART) involves combinations of small-molecule drugs that act at various cellular sites, protecting uninfected target CD4<sup>+</sup> T cells from HIV infection by inhibiting viral entry or post-entry viral function. But this treatment does not affect the production of virus by infected cells, nor facilitate the clearance of virus particles. Persistent virus-producing cells and circulating virus can reignite progressive infection if ART is discontinued. Free virus can also activate cells, contributing to a chronic inflammatory environment. **b**, Monoclonal antibodies (mAbs) that bind to and inactivate viral envelope proteins effectively neutralize the virus and therefore also prevent the spread of infection. These may have the added advantage of facilitating the clearance of virus particles and killing virus-producing cells by antibody-dependent mechanisms, in particular antibody-dependent cell-mediated cytotoxicity (ADCC), thereby potentially diminishing the viral reservoir and preventing local immune activation.

glycans) profoundly suppressed viraemia, often to below the level of detection. This suppression persisted for weeks, as long as mAb levels were maintained. Furthermore, mutations associated with resistance to the neutralization activity of the administered mAbs were not generally observed in the virus that rebounded after mAb levels waned. Remarkably, Barouch and colleagues show that one antibody, the extraordinarily potent N332-glycan-dependent PGT121 mAb<sup>5</sup>, was able to provide prolonged viral suppression when used alone.

Given that current antiretroviral therapy (ART), which is based on small molecules that target different phases of the HIV life cycle, are effective and can be given orally, whereas mAbs need to be delivered by injection, why is this approach of interest? The answer lies in the different mechanisms of viral suppression by conventional ART versus neutralizing mAbs. Current ART drugs prevent the virus from spreading from an infected cell to the next round of target cells, but they do not directly kill infected cells nor inhibit the cells' production of virus particles<sup>6</sup> (Fig. 1a). In the absence of viral spread, virus production declines as infected cells die by virus- or immune-mediated killing or by programmed cell death. But some infected cells persist and produce virus particles indefinitely. This rare but enduring reservoir of virus-producing cells is the source of viral rebound when people stop ART, and

probably contributes to persistent inflammation and immune dysfunction (and associated morbidities) in individuals on ART<sup>7,8</sup>.

Treatment with neutralizing mAbs would, like conventional ART, prevent viral spread, and this undoubtedly accounts for much of the antibodies' observed activity in these two studies. However, mAbs have intrinsic effector functions that would not only facilitate rapid clearance of virus particles produced by residually infected cells, but might also accelerate the destruction of virus-producing cells by antibody-dependent cell-mediated cytotoxicity, complement lysis or phagocytosis (Fig. 1b). Although these processes were not directly demonstrated in either of the new studies, Barouch *et al.* show accelerated viral clearance in mAb-treated macaques compared with that seen in humans given conventional ART. They also provide evidence for declining levels of cell-associated HIV DNA (a measurement of the number of infected cells) and improved function of SHIV-specific CD8<sup>+</sup> T cells (antiviral immune cells) in mAb-treated animals. Perhaps more interestingly, Barouch and colleagues report that the viral load set points — the stable level of viraemia reached once a balance is established between viral replication and the immune system — of the animals after therapy cessation were consistently lower than before treatment, suggesting that mAb treatment enhanced

the antiviral immune response.

There are several caveats to this exciting work that deserve mention. One is the substantial animal-to-animal variability in pharmacokinetic outcomes, which would have to be addressed if long-term administration of these antibodies were pursued in people. Another is the fact that the high mutability of HIV means that slightly different envelope proteins will be expressed by virus strains both within an individual and across populations, making it likely that some people will harbour viruses that are resistant to one or more of the therapeutic mAbs. In addition, whether the mAbs can cross the blood–brain barrier and suppress viral replication in the central nervous system remains an open question. Finally, the extent to which viral envelope protein is exposed on the surface of long-lived infected cells is unknown. Additional treatment with virus-activating agents will probably be needed for antibody-mediated destruction to have a meaningful impact on this reservoir.

Despite these limitations, the findings of these two papers could revolutionize efforts to cure HIV. The combination of conventional ART and neutralizing mAb therapy might reduce viral replication more effectively than conventional ART alone, mediate destruction of virus-producing cells arising from latent infection, diminish deleterious, generalized immune activation, and facilitate the efficacy of T-cell-targeted therapeutic vaccines intended to destroy latent cells<sup>8</sup>. At the very least, these results will catalyse collaborations between the massive teams of experts who have for decades worked on HIV prevention and treatment in separate venues. As previously emphasized<sup>9,10</sup>, true multidisciplinary science will be needed if we are to generate a cure for HIV. ■

**Louis J. Picker** is at the Vaccine and Gene Therapy Institute, Oregon Health & Science University, Beaverton, Oregon 97006, USA.

**Steven G. Deeks** is in the Department of Medicine, University of California, San Francisco, San Francisco, California 94110, USA.

e-mails: pickerl@ohsu.edu;  
sdeeks@php.uscf.edu

- Mascola, J. R. & Haynes, B. F. *Immunol. Rev.* **254**, 225–244 (2013).
- Klein, F. *et al. Science* **341**, 1199–1204 (2013).
- Barouch, D. H. *et al. Science* **341**, 224–228 (2013).
- Shingai, M. *et al. Nature* **503**, 277–280 (2013).
- Moldt, B. *et al. Proc. Natl Acad. Sci. USA* **109**, 18921–18925 (2012).
- Volberding, P. A. & Deeks, S. G. *Lancet* **376**, 49–62 (2010).
- Palmer, S. *et al. Proc. Natl Acad. Sci. USA* **105**, 3879–3884 (2008).
- Deeks, S. G. *et al. Nature Rev. Immunol.* **12**, 607–614 (2012).
- Deeks, S., Drosten, C., Picker, L., Subbarao, K. & Suzich, J. *Nature Med.* **19**, 30–34 (2013).
- Richman, D. D. *et al. Science* **323**, 1304–1307 (2009).

This article was published online on 30 October 2013.



# Sound and heat revolutions in phononics

Martin Maldovan<sup>1,2</sup>

**The phonon is the physical particle representing mechanical vibration and is responsible for the transmission of everyday sound and heat. Understanding and controlling the phononic properties of materials provides opportunities to thermally insulate buildings, reduce environmental noise, transform waste heat into electricity and develop earthquake protection. Here I review recent progress and the development of new ideas and devices that make use of phononic properties to control both sound and heat. Advances in sonic and thermal diodes, optomechanical crystals, acoustic and thermal cloaking, hypersonic phononic crystals, thermoelectrics, and thermocrystals herald the next technological revolution in phononics.**

In recent decades, major technological revolutions have transformed our society and daily lives. Their remarkable innovations have been based primarily on our improved ability to manipulate two particles: electrons and photons. In particular, the control of electrons in semiconductor materials has generated fundamental changes, with laptops, mobile telephones and digital cameras now products that seem always to have existed. Analogously, the development of materials and devices with which to control photons has generated major changes in society, such as wireless communication and the use of optical fibres and microwaves. The successful management of the electromagnetic spectrum is demonstrated by the wide range of frequencies controlled in electromagnetic devices, which extends over 14 orders of magnitude, ranging from positron emission tomography (PET) scanning at frequency  $f \approx 10^{20}$  Hz to amplitude-modulation (AM) radios at  $f \approx 10^6$  Hz.

Besides electrons and photons, another everyday particle is the phonon, which is responsible for the transmission of sound and heat. Given the many applications of our remarkable success in managing electrons and photons, it would be valuable to achieve a similar degree of control over the particle that accounts for both sound and heat (Fig. 1). Although some sonic and thermal devices and materials, such as medical ultrasound imaging machines and thermal insulation materials, are well known, further applications for phononic devices are developing fast, from thin acoustic metamaterials that can soundproof rooms to enhanced thermoelectric devices that can use our bodies' waste heat to power portable electronic devices.

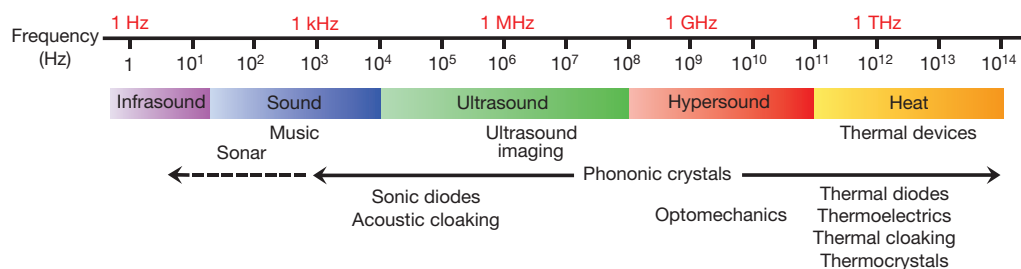
In this review, I provide insights on newly developed phononic materials and devices that have recently been shown to have remarkable physical properties—making it possible to manipulate sound and heat with unprecedented precision. I begin with a classical description of small-scale phononic crystals, followed by recent progress on unidirectional sound transmission in acoustic diodes—analogue to electronic diodes—and new developments on acoustic cloaking using metamaterials. I examine

novel interactions between sound and light in resonant cavities, and discuss new advances in thermal diodes and thermal metamaterials. I also provide an overview on thermoelectric materials (which convert heat into electricity) and discuss novel thermocrystals, which treat heat like sound and bring phonon management to the boundary zone between sound and heat. These recent ideas and concepts have improved our ability to manage the phononic spectrum, providing exciting opportunities to develop a large variety of devices that can control sound and heat.

## Hypersound and heat control by phononic crystals

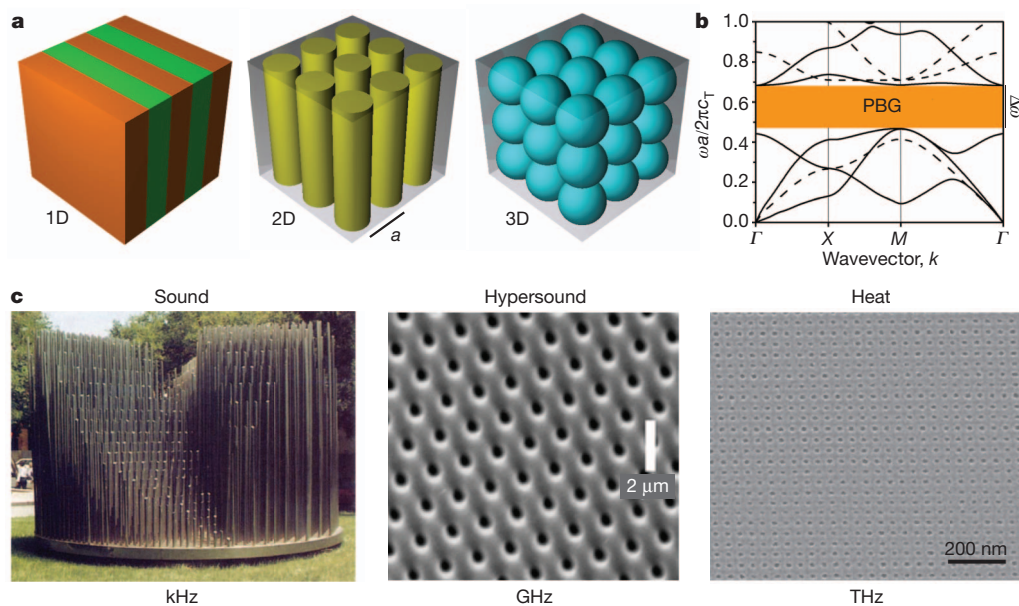
Sound and heat can both be described as mechanical vibrations transmitted through the atomic lattice. One difference between them, however, is that most sound waves oscillate at low frequencies (kilohertz) and propagate over large distances, whereas most heat vibrations oscillate at high frequencies (terahertz) and travel small distances. These different features lead scientists and engineers to employ different strategies to control sound and heat propagation. Essentially, macroscale and microstructured materials are able to manipulate sound and hypersound (very high-frequency sound, with  $f \approx 1$  GHz) frequencies, whereas to control heat, nanostructures are generally required. The development of sonic and thermal devices thus requires the design, fabrication and characterization of composite materials ranging from the centimetre scale to the nanometre scale.

Over the past two decades, the propagation of sound waves with frequencies in the range from kilohertz to megahertz has been efficiently controlled by phononic crystals—artificial periodic structures made of two elastic materials<sup>1,2</sup> (Fig. 2). Although these structures were initially designed to control sound, in recent years small-scale phononic crystals have been successfully used to control hypersound and heat. In phononic crystals, mechanical waves with frequencies within a specific range are not allowed to propagate within the periodic structure. This forbidden frequency range—the 'phononic bandgap' (PBG)—allows sound to



**Figure 1 | The phononic spectrum.**

<sup>1</sup>Department of Materials Science and Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. <sup>2</sup>School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, North Avenue, Atlanta, Georgia 30332, USA.



**Figure 2 | Phononic crystals.** **a**, 1D, 2D and 3D phononic crystals made of two different elastic materials arranged periodically. Different colours represent materials with different elastic properties. **b**, An example of a phononic band diagram  $\omega = \omega(k)$  for a two-dimensional phononic crystal, in which non-dimensional frequencies  $\omega a/2\pi c_T$  (with  $c_T$  a transverse velocity) are plotted versus the wavevector  $k$  along the  $\Gamma$ -X-M- $\Gamma$  path in the square Brillouin zone. The range of forbidden frequencies, or PBG, is shown in orange. **c**, 2D phononic crystals with periodicities  $a$  in the centimetre range (left), the micrometre range (middle) and the tens-of-nanometres range (right) can be used to control sound, hypersound and heat, respectively. Images are taken from fig. 1 of ref. 3 (left image, with permission), fig. 1a of ref. 8 (middle image) and fig. 1b of ref. 11 (right image, with permission).

be controlled in many useful ways in structures that can act as sonic filters, isolators, waveguides or resonant cavities. Owing to wave interference effects, PBGs occur for phonon wavelengths  $\lambda$  that are comparable to the structure periodicity  $a$  (Fig. 2). For experimental convenience, a large number of phononic crystals have been fabricated at macroscopic scales ( $a \approx 10$  cm–1 mm) to control sound of frequency  $f \approx 10^3$  Hz and ultrasound of  $f \approx 10^6$  Hz (refs 3–7), finding promising applications in acoustics, medical diagnosis and remote sensing. In recent years, however, the challenge has been to fabricate phononic crystals having small periodicities  $a < 1$   $\mu$ m, to control high-frequency ( $f \approx 10^9$ – $10^{12}$  Hz) phonons<sup>8–11</sup>. Given that the periodicity of small-scale ( $a \approx 1$   $\mu$ m) phononic crystals can be comparable to the wavelengths of light, these crystals can provide bandgaps for both hypersound and light, suggesting new mechanisms with which to enhance phonon–photon interactions<sup>12</sup>. By additionally reducing the periodicity to the nanometre scale, the applications of small-scale phononic crystals can be extended from sound to heat, because PBGs could in principle exist in the terahertz regime. As shown later, however, this reasoning requires a more precise analysis owing to the different characteristics of sound and heat transfer.

A small-scale phononic crystal for hypersound frequencies  $f \approx 1$  GHz has been fabricated using interference lithography<sup>8</sup>. This technique provides single-crystalline periodic structures by transferring a laser intensity pattern into a photosensitive material. The phononic crystal consists of a triangular array of air holes in an epoxy matrix, with lattice periodicity  $a = 1.36$   $\mu$ m (Fig. 2c). To characterize the phonon modes, Brillouin light scattering is employed, which provides complete phononic band diagrams for the dispersion relation for phonon modes propagating within the crystal (Fig. 2). Other small-scale phononic crystals were constructed by self-assembling single-crystalline colloidal films of polystyrene spheres into a face-centred-cubic lattice<sup>9</sup>. The small size of the spheres used ( $d = 256$  nm) guarantees operation in the hypersonic frequency range. By employing Brillouin light scattering to map the dispersion relation of longitudinal phonons, a hypersonic (gigahertz frequency) PBG was measured. The significance of these small-scale architectures is that, in addition to providing simultaneous gaps for hypersound and light, they greatly extend the use of phononic crystals from sonic and ultrasonic frequencies to the hypersonic regime<sup>8–10</sup>. This increases the range of phononic frequencies that can be controlled by three orders of magnitude.

More recently, a state-of-the-art nanometre-scale phononic crystal was fabricated by transferring the patterns of two intersecting arrays of platinum nanowires into a silicon epilayer (Fig. 2c)<sup>11</sup>. The phononic

‘nanomesh’ consists of a square-lattice silicon matrix of air holes, with  $a = 34$  nm. The key feature of this structure is that it reduces the thermal conductivity of bulk silicon by almost two orders of magnitude. Heat in silicon is mostly carried by phonons with  $10^{12}$  Hz frequencies, so this demonstrates that reducing the size of phononic crystals to the nanometre scale can modify and control the propagation of high-frequency phonons, extending the applications of phononic crystals from sound to heat<sup>11</sup>. As I explain below, however, the mechanism behind the reduction of the thermal conductivity is not based on coherent interference and PBGs—as in classical phononic crystals—but is mainly due to the diffuse scattering of phonons at interfaces.

With the recent progress in synthesis and characterization of small-scale materials, it is now possible to fabricate phononic crystals with periodicities ranging from centimetres to nanometres<sup>3–11</sup>. These advances have significantly increased the ability to manage the phonon spectrum, where the frequencies that can now be controlled extend from  $10^3$  Hz to  $10^{12}$  Hz, a range of nine orders of magnitude. As I show in this review, the wide range of control over phonon frequencies—comparable to the wide range of control over electromagnetic frequencies—allows the realization of many exciting devices that can manipulate phonons at all relevant frequencies: sound, ultrasound, hypersound and heat.

## Acoustic diodes

In electronics, a diode is a well-known device that allows electrical current to pass in one direction but not in the opposite direction. This fundamental principle—which led to a revolution in electronics—has recently been demonstrated for sound waves<sup>13–15</sup>. The development of the acoustic counterpart of the electronic diode opens up exciting opportunities for novel applications in different research areas ranging from biomedical ultrasound imaging to environmental noise reduction.

A sonic diode showing unidirectional transmission was constructed using a one-dimensional (1D) phononic crystal made of alternating layers of glass and water coupled to a nonlinear acoustic medium<sup>13,14</sup> (Fig. 3a). A sound wave incident from the left, with frequency  $\omega$  inside the PBG, is partially converted within the nonlinear medium to a secondary wave with frequency  $2\omega$ . Because  $\omega$  lies in the PBG, the original wave is reflected backwards, but the secondary wave with frequency  $2\omega$  outside the PBG can freely pass through the structure (Fig. 3a). On the other hand, if the sound wave is incident from the right, the wave is completely reflected in the backward direction because  $\omega$  lies inside the PBG. As a result, a sound wave with frequency  $\omega$  can pass through the



system only along the forward direction, providing unidirectional sound transmission. In experiments<sup>14</sup>, the layers were about 1 mm thick, allowing the diode to operate at ultrasound frequencies in the megahertz range.

An even more interesting approach using linear materials has been introduced by means of a two-dimensional (2D) phononic crystal made of a square lattice of steel rods in contact with a diffraction structure<sup>16</sup> (Fig. 3b). The critical feature in this diode is that one of the external surfaces is 'smooth' while the other is 'rough'. Waves incident on the smooth surface (reverse direction) with frequencies  $\omega$  inside the PBG ( $\sim 10$  kHz) are reflected backwards because they fall into the forbidden frequency range. In contrast, waves incident on the rough surface (forward direction) can pass through partially. Because the outgoing waves in the forward direction are not parallel to the incident waves, the acoustic energy is converted through high-order diffractions to other modes with different spatial frequencies to overcome the barrier imposed by the PBG<sup>16</sup>. The main advantage of the linear acoustic diode is that sound waves do not change their frequencies as they pass through the structure. Also, its easy fabrication method opens up opportunities for on-chip planar acoustic circuits with applications in logical sound processing.

Sound rectification has also been demonstrated in 1D granular crystals<sup>17</sup> consisting of a chain of steel spheres interacting nonlinearly (Fig. 3c). This system allows wave propagation for frequencies lower than a certain cut-off  $\omega_C$ . When a defect sphere is added, a localized mode with frequency  $\omega_D > \omega_C$  is created around the defect. An incident wave with  $\omega > \omega_C$ , and below  $\omega_D$ , along the reverse direction (right to left) cannot propagate through the crystal because of the PBG. However, if moving along the forward direction (left to right), a wave with frequency  $\omega$ —having low amplitude—does not pass through but excites the vibrational mode localized around the defect sphere. If the wave's amplitude is increased, its acoustic energy is converted to modes with frequencies outside the gap ( $\omega < \omega_C$ ) that can freely pass through the structure, generating unidirectional transmission.

The experimental realizations of acoustic diodes, providing unidirectional sound transmission, are major steps towards complex acoustic devices that could greatly increase the efficiency of ultrasound medical

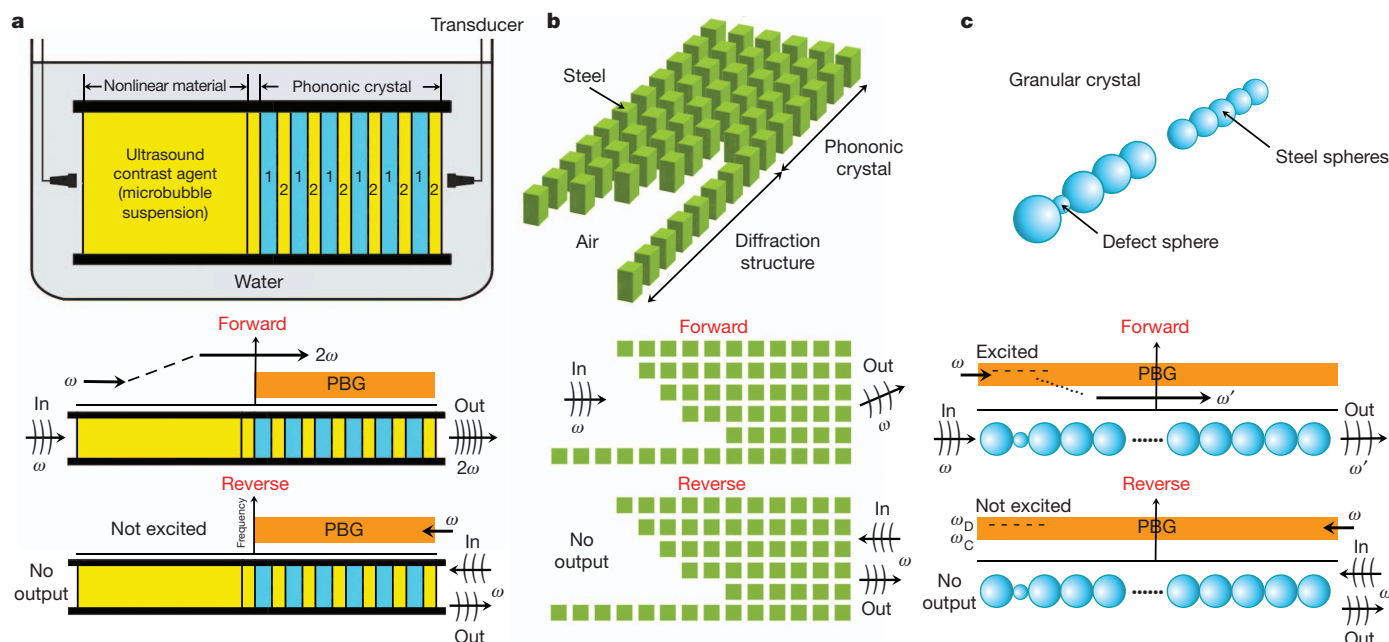
systems, non-destructive testing and environmental noise control. Like its electric counterpart, the sonic diode may produce unexpected consequences, in acoustics rather than electronics.

## Acoustic invisibility cloaks by metamaterials

Another exciting and novel opportunity to control sound waves involves the design and fabrication of acoustic cloaking shells. The basic idea is to use a shell to guide acoustic waves around a certain object such that the object inside the shell becomes 'invisible' (undetectable by sound waves) (Fig. 4a). Although cloaking devices were initially introduced for electromagnetic waves<sup>18,19</sup>, these principles have recently been extended to sound waves<sup>20–33</sup>. Essentially, to obtain a cloaking device for acoustic waves, the cloaking shell surrounding the object must be made of an engineered artificial material, known as acoustic metamaterial, in which elastic properties such as mass density and bulk modulus are anisotropic and spatially dependent.

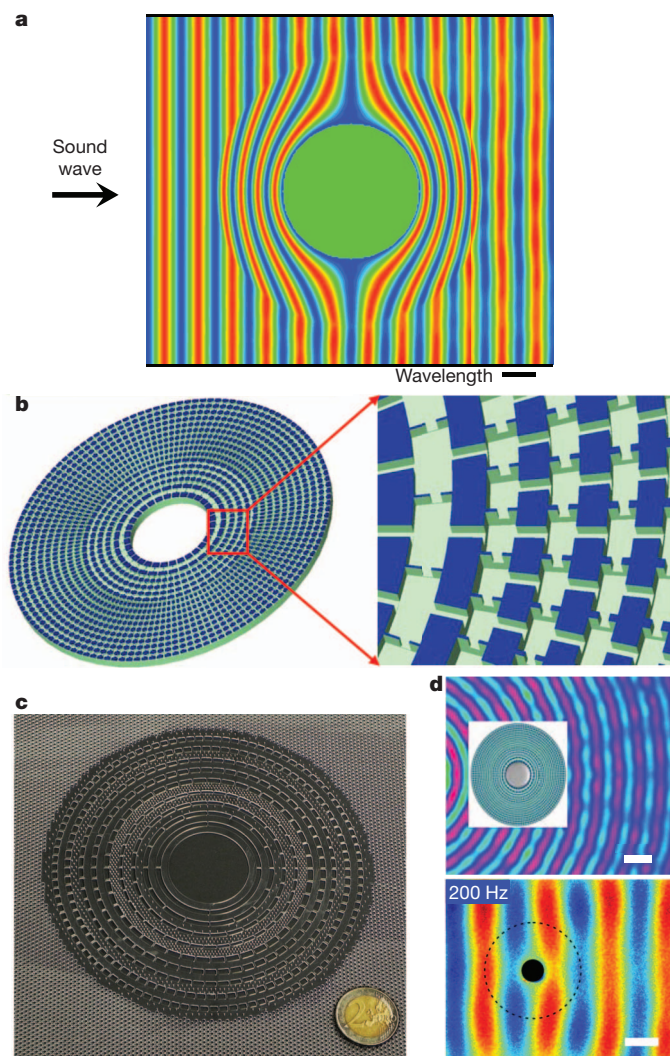
An acoustic cloaking shell was introduced by considering the similarity between the 2D electromagnetic and acoustic wave equations<sup>21</sup>. The cylindrical shell is made of a metamaterial fluid with inhomogeneous bulk modulus and an anisotropic and inhomogeneous mass density that redirects pressure waves around the object (Fig. 4a). Cloaking effects have been extended to three dimensions by noting the isomorphism between the acoustic wave equation and the electrical conductivity equation<sup>22</sup>. Such three-dimensional (3D) cloaks are also obtained by considering zero scattering effects in the acoustic wave equation<sup>23</sup>. Because the 2D and 3D cloaking shells require complex materials, it is challenging to implement such metamaterials in experiments. Owing to this difficulty, additional theoretical approaches have been developed in which layered cylindrical cloaking shells are considered to obtain the required values for mass density and bulk modulus<sup>24–26</sup>.

To overcome the difficulty associated with material requirements, an innovative approach based on the analogy between acoustic elements and electronic circuits has been employed to design and fabricate a 2D cylindrical cloak (Fig. 4b)<sup>27,28</sup>. The basic idea is to convert the acoustic material to an analogous network of inductors  $L$  and capacitors  $C$ , where



**Figure 3 | Acoustic diodes.** **a**, An acoustic diode made of a 1D phononic crystal (alternating layers of glass and water) coupled to a nonlinear acoustic medium<sup>14</sup>. A sound wave with frequency  $\omega$  within the PBG can evade the PBG—and pass through the system—only from left to right (forward direction), owing to the change of frequency  $\omega \rightarrow 2\omega$  in the nonlinear medium. **b**, A linear acoustic diode consisting of a 2D phononic crystal (square steel rods with  $d = 4$  mm in a square lattice with  $a = 7$  mm) and a diffraction structure<sup>16</sup>.

Waves with frequency  $\omega$  within the PBG cannot pass through the structure if they are incident on the phononic crystal (reverse direction), but they can propagate if they are incident on the diffraction structure (forward direction). **c**, Sound rectification in 1D granular crystals (steel spheres with radius 9.53 mm and mass 28.84 g)<sup>17</sup>. A defect sphere creates sound rectification by introducing localized states with frequencies  $\omega_D$  that allow incident waves to evade the PBG (and pass through) by changing frequencies from  $\omega$  to  $\omega'$  (forward direction).



**Figure 4 | Acoustic cloaking.** **a**, A sound wave incident on a rigid cylinder surrounded by a cloaking shell. The wave passes through the shell and propagates as it would have done in the absence of an obstacle. Reproduced from fig. 2 of ref. 25 with permission. **b**, A cylindrical acoustic cloak for ultrasound waves in water. Reproduced from fig. 2a of ref. 27 with permission. **c**, Cloaking shell for elastic waves made of concentric cylinders of PVC and PDMS. The coin diameter is 25.75 mm. Reproduced from fig. 2 of ref. 31 with permission. **d**, The measured pressure from a 60-kHz point source for a steel cylinder (grey) surrounded by the cloaking shell (light-blue coating around the cylinder) from **b**, where waves pass largely unperturbed (top). Scale bar, 2.7 cm. Elastic waves propagating through a cylinder (black) with the cloaking shell (dotted circular line) from **c**. The shell avoids disruption of the elastic waves (bottom). Scale bar, 5 cm. Reproduced from figs 3d and 3b of refs 27 and 31 with permission.

the motion of the fluid is equivalent to the electrical current, and the mass density and bulk modulus are related to the capacitance  $C$  and inductance  $L$ . Using this ‘transmission line’ approach, an aluminium plate having an array of cavities (acting as acoustic capacitors) connected by channels (acting as acoustic inductors) was designed to perform as an acoustic cloak for ultrasound waves (Fig. 4b). The aluminium-based cloak was placed in a water tank (with an object within the shell) to compare ultrasound waves around the object with and without the cloak. When the object is surrounded by the metamaterial cloak, the incoming wave is restored behind the object with very little distortion, making the cloak and the object ‘invisible’ to ultrasound waves in water<sup>27</sup>. This acoustic cloak—fabricated in the millimetre length scale—operates over a wide range of ultrasound frequencies (40–80 kHz). A previous cloak, fabricated at the centimetre scale, but designed for surface waves in a liquid material, has also provided cloaking effects in the range of 10-Hz frequencies<sup>29</sup>.

Besides full cloaks in water, an acoustic cloak that works in air, and for audible frequencies 1–4 kHz, has been designed, fabricated and characterized using acoustic metamaterials<sup>30</sup>. This cloak, however, is based on a different technique known as ‘ground cloaking’, because the object to be made ‘invisible’ is located on a reflecting surface. The object on the surface is covered with a coating whose metamaterial properties are given by coordinate transformation acoustic theory. The function of the coating is to make the object invisible. The metamaterial coating is made of a stack of planar perforated plastic plates, in which the size (millimetres) and shapes of the perforations are used to design the acoustic properties for cloaking. Using a 3-kHz speaker as a sound source, it is shown that the reflected acoustic field from the ground surface alone and that from the ground surface and the coated object are essentially identical, thus demonstrating a new way of cloaking objects from sound waves in air<sup>30</sup>.

A cloaking shell for elastic waves<sup>31</sup>, fabricated using a structured polymer plate, demonstrates that theory and experiments on acoustic cloaking can be expanded from fluids to solid materials (Fig. 4c)<sup>31–33</sup>. The solid cloaking shell consists of 20 concentric rings, each a tailored composite of polyvinyl chloride (PVC) and polydimethylsiloxane (PDMS) that allowed the Young’s modulus within the shell to be specified and cloaking effects to be obtained. To show the cloak in operation, elastic sound waves were guided through the shell (and around the object to be concealed) and their movement was recorded using a stroboscopic imaging technique. For frequencies of 200–400 Hz, the elastic waves can pass through the cloaking shell and propagate as they would have done without an obstacle—thus demonstrating acoustic cloaking in solid materials (Fig. 4d). This cloak is based on a theoretical proposal that modifies the technique of coordinate transformation of Maxwell’s equations to create cloaking shells for elastic flexural waves in solids<sup>32</sup>. Recently, a theoretical approach for the control of coupled pressure and shear waves in elastic solids has also been introduced<sup>33</sup>.

The remarkable progress made in recent years in acoustic cloaking is evident from the large number of devices designed to hide objects from sound waves, which include air, water and solid-material platforms. In analogy with their electromagnetic counterpart, acoustic cloaking and metamaterials open up new ways to manipulate sound. Such metamaterials can be used to soundproof studios and buildings, design concert halls, control environmental noise, make ships invisible to sonar waves, or even design shields against seismic waves. Metamaterial concepts and devices bring together different research areas, including cosmology, physics, materials science and civil engineering.

### Enhancing the interaction between sound and light

Interaction between phonons and photons is well known in studies of Raman spectroscopy and Brillouin scattering processes. Recently, a mechanism involving the simultaneous localization of mechanical and optical waves by using acoustic and optical cavities has been introduced greatly to enhance sound and light interactions. This raises the prospect of a new class of acousto-optical crystal that can integrate the combined management of phonons and photons<sup>12,34–59</sup>.

A 1D crystal showing simultaneous localization of sound and light by resonant cavities was introduced by constructing an acoustic cavity within another optical cavity<sup>34,35</sup> (Fig. 5a). The optical cavity is made of  $\text{Al}_x\text{Ga}_{1-x}\text{As}$  and AlAs layers grown by molecular beam epitaxy, within which a spatially localized optical mode (red) can be created by introducing a defect layer. Similarly, the acoustic cavity is constructed by GaAs and AlAs layers, with a defect  $\text{Al}_x\text{Ga}_{1-x}\text{As}$  layer that localizes sound modes (blue). The key feature of the system is that, owing to their relatively large wavelengths, photons see the acoustic layered structure as a homogeneous defect layer. In particular, when an incident laser photon excites the localized mode in the optical cavity, this mode splits photoelastically into a localized phonon mode in the acoustic cavity and a localized Stokes mode in the optical cavity. These localized phonons and photons leave the cavities by tunnelling through the layers. Measurements of the light spectrum for the outgoing photons clearly show the energy shift caused by these novel confined acousto-optical interactions<sup>34,35</sup>.



A different approach, in which sound and light are both localized in the same cavity, considered the use of periodic structures acting as both phononic and photonic crystals (Fig. 5b)<sup>12</sup>. In analogy to phononic crystals, periodic dielectric structures such as photonic crystals can precisely control light propagation<sup>36–39</sup>. An effective mechanism with which to create non-propagating localized states is the introduction of a defect in a periodic structure having a frequency bandgap<sup>38,39</sup>. Although extensive work has been focused on separate architectures for phononics and photonics, the existence of a PBG does not imply that the structure has a photonic bandgap and vice versa. These dual-bandgap (or ‘deaf and blind’) structures have been demonstrated by considering an array of air cylinders in a silicon matrix (Fig. 5b)<sup>12,39</sup>. By taking advantage of simultaneous gaps, a defect is introduced to create a single cavity that localizes both phonons and photons. In particular, it was indicated that if the structure periodicity is  $a \approx 150$  nm, the cavity can co-localize 500-THz photons and 20-GHz phonons<sup>12</sup>. The simultaneous confinement of

phonons and photons in a single cavity is suggested greatly to increase acousto-optical interactions, leading to a different class of planar acousto-optical devices that can integrate both light and sound management<sup>12</sup>.

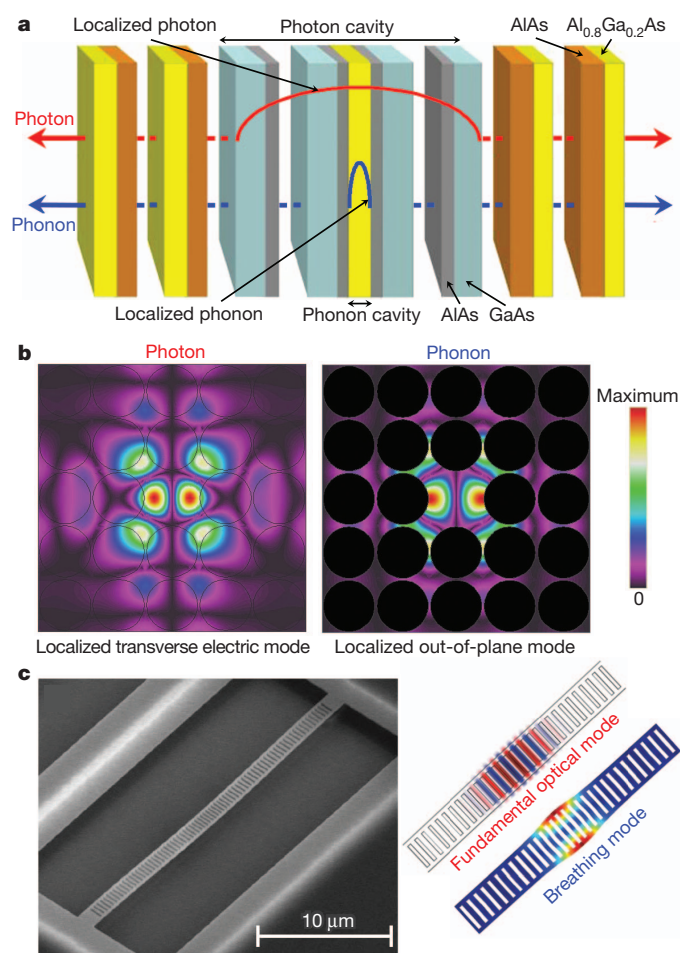
More recently, co-localization and interaction of 200-THz photons and 2-GHz phonons has been demonstrated<sup>40</sup> by the design, fabrication and characterization of a 1D silicon nanobeam with rectangular air holes (Fig. 5c). To co-localize optical and mechanical modes at the centre of the structure, a ‘defect’ is introduced by quadratically—and symmetrically—decreasing the periodicity from the nanobeam centre. Electric and mechanical field distributions for co-localized optical and mechanical modes are shown in Fig. 5c. In this system, the optical modes are excited by using a tapered optical fibre that also collects the transmitted light. Mechanical modes that can couple to optical modes are classified (according to their increasing frequencies) as ‘pinch’, ‘accordion’ and ‘breathing’. By measuring their radio-frequency spectrum, the character of these modes is described by evaluating the phase and amplitude change of the transmitted light. Importantly, the demonstration of phonon–photon coupling in a planar ‘optomechanical’ crystal allows for new methods of probing, stimulating, and using mechanical and optical interactions on a chip-scale platform<sup>40</sup>.

The prospects of manipulating sound and light by co-localization in periodic crystals have stimulated significant research in recent years. This includes 1D vertical-cavity crystals<sup>41,42</sup>, 2D and quasi-2D planar crystals<sup>43–53</sup> and 3D crystals<sup>54,55</sup>, involving exciting phenomena such as photonic gaps, optomechanical coupling, laser cooling, light modulation, quantum motion and phonon laser action<sup>41–55</sup>. In addition, acousto-optical studies on crystal fibres (which have a solid silica central core surrounded by a lattice of hollow channels extending axially along its length) show interesting phenomena caused by the strong confinement of light and acoustic vibrations<sup>56–59</sup>. These effects include multi-peak Brillouin spectra with frequency shifts in the 10-GHz range, as well as Raman-like nonlinear forward scattering of light with gigahertz phonons<sup>58,59</sup>. These results show that separate research areas such as those involving sound control in phononic crystals and light management in photonic crystals can be successfully merged into a field of dual phononic–photonic crystal materials.

## Thermal diodes

The experimental realization of materials and devices that can provide unidirectional transmission of heat is more difficult than in the case of acoustic diodes. This is because heat is carried by a broad spectrum of high-frequency (terahertz) phonons, which are difficult to control. Despite this, novel theoretical and experimental concepts have been introduced whereby heat can flow only in a particular direction<sup>60</sup>. A theoretical thermal diode was proposed by means of two coupled one-dimensional lattices with particles interacting with each other through nonlinear sinusoidal potentials (Fig. 6a). In the forward direction ( $T_L > T_R$ ), the broad spectrum of one chain overlaps with that of the other chain, allowing heat to flow through the structure. In the reverse direction ( $T_L < T_R$ ), however, the vibrational frequencies of the lattices do not match and heat cannot flow through. This match–mismatch of the vibrational modes gives rise to the unidirectional transmission of thermal energy. Thermal rectification effects are achieved by carefully considering different strengths in the potentials of the two segments of the lattice<sup>60</sup>. This theoretical model is a simplified version of a previously proposed diode involving a three-segment 1D nonlinear chain<sup>61</sup>. In recent years, the thermal diode concept has been expanded to include the thermal counterparts of the components of digital electronic circuits—the thermal transistor and thermal logic gates<sup>62,63</sup>. However, important material challenges must still be overcome before their practical realization and use as thermal devices.

In experiments, thermal diodes were fabricated using carbon nanotubes and boron nitride nanotubes<sup>64</sup>. To investigate asymmetric thermal motion, amorphous  $C_9H_{16}Pt$  is deposited non-uniformly along the length of the tubes (Fig. 6b). When the heat flows from the high-mass region (left) to the low-mass region (right), the measured thermal conductance is higher. Specifically, for carbon nanotubes and boron nitride nanotubes respectively, the conductance along the nanotube axis is 2%



**Figure 5 | Enhancing sound–light interaction.** **a**, An array of semiconductor layers creates an acoustic cavity within an optical cavity. A light beam entering the structure excites a localized light mode (red) within the optical cavity and a localized sound mode (blue) within the acoustic cavity. Light and sound escape by tunnelling—allowing their interaction to be measured. **b**, Periodic structures having dual photonic–phononic bandgaps can localize both sound and light in the same cavity. Transverse electric (TE) photons are localized within a structure made of air cylinders in silicon by creating a cavity (the missing cylinder) (left). Out-of-plane phonons are localized in the same dual photonic–phononic cavity (right). Co-localization can greatly enhance acousto-optical interactions. Images taken from fig. 3 of ref. 12. **c**, Co-localization and interaction of photons and phonons in a nanobeam with a central defect. Scanning electron microscope image for the nanobeam device (left). Optical and mechanical field distributions for co-localized and coupled opto-mechanical modes (right). Reproduced from figs 1 and 2a of ref. 40 with permission.

and 7% larger than that in the reverse direction. To interpret these results, it was suggested that boron nitride nanotubes show larger thermal rectification effects owing to their strong ionic bonds, which favour nonlinear effects. This mechanism has subsequently been theoretically described by considering a 1D anharmonic lattice with a mass gradient. In agreement with experiments, this system provides a larger flow of heat when the heavy-mass end is set at a higher temperature<sup>65</sup>. The major impact of the electrical diode in electronic circuits suggests that highly efficient thermal diodes could have a similar influence in a large number of systems that require precise heat control, such as nanoelectronic devices and energy-saving buildings.

### Thermal metamaterials and heat cloaking

Another interesting idea is to apply the metamaterial methods that are used for sound manipulation to the management of heat conduction. This concept has recently been proposed by realizing that mathematical techniques to transform Maxwell's equations<sup>18,19</sup> can be adapted to the Fourier heat equation<sup>66–70</sup>. Controlling heat conduction has always been a difficult scientific challenge. Using newly designed thermal metamaterials<sup>66–70</sup>, however, several striking effects on diffusive heat transfer have been implemented, offering a new way to manipulate heat flow (Fig. 7).

Thermal metamaterials were introduced by considering a cylindrical shell that, under a thermal gradient, maintains the inner region at constant temperature, that is, the core is 'invisible' to the thermal gradient<sup>19,66</sup>. Interestingly, it was also suggested that an inverse thermal flow (or apparent negative thermal conductivity) might be achieved by changing the geometrical shape of the shell<sup>66</sup>. These innovative theoretical concepts have been realized in experiments by constructing a thermal shield made of alternating concentric cylinders of latex rubber and a silicone elastomer, and placing the shield within an agar–water matrix (Fig. 7a)<sup>69</sup>. Under a thermal gradient, a constant temperature profile is obtained in the shield's inner region while the temperature profile outside is almost undistorted. On the other hand, by intentionally arranging the materials along the radial direction, the shell can act as a thermal concentrator, with the thermal energy considerably enhanced within the shell and remaining uniform (Fig. 7b). Even more surprisingly, for a spiral arrangement of the layers in the shell, heat rotates within the composite shell in a spiral manner such that heat flux (and the thermal gradient) changes its sign

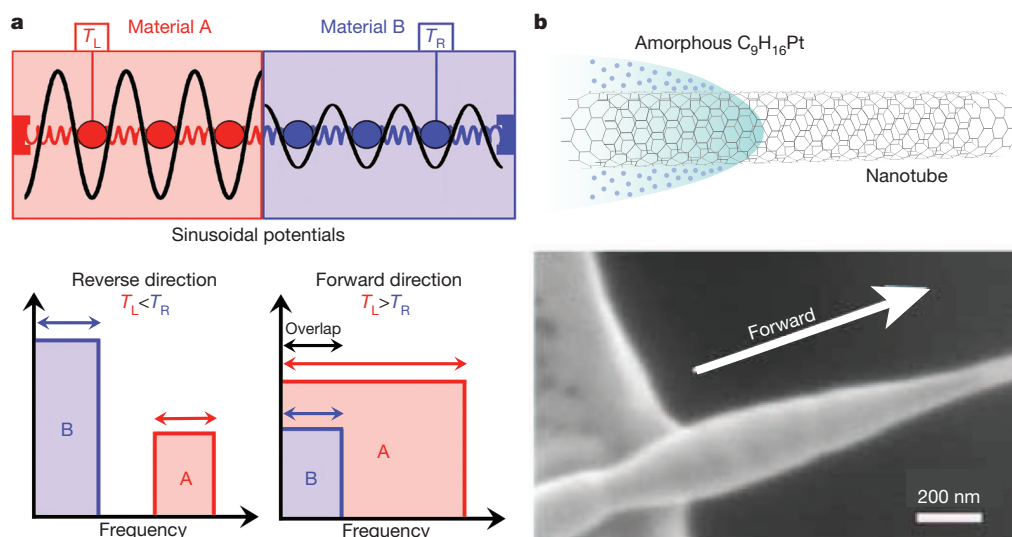
within the inner region of the shell—creating an apparent negative thermal conductivity (Fig. 7c)<sup>69</sup>.

Thus heat management using metamaterials brings exciting new techniques with which to manipulate thermal energy. In analogy to optical and acoustical metamaterials, these offer unprecedented control of heat conduction.

### Nanoscale heat transport and thermoelectrics

An additional important aspect of heat management is the need to reduce the amount of heat transport to ultralow levels. Materials with low thermal conductivities strongly affect the efficiency of thermoelectrics—which convert waste heat to electricity and provide an alternative source of energy for the future. In recent years, different strategies have been employed to enhance the efficiency of thermoelectrics, which is determined by the figure of merit  $ZT = S^2\sigma T/\kappa$ , where  $T$  is the temperature,  $S$  is the Seebeck coefficient,  $\sigma$  is the electrical conductivity and  $\kappa$  is the thermal conductivity. One approach involves the manipulation of the electronic properties—using quantum wells and wires—such that  $S$  and  $\sigma$  (and thus  $ZT$ ) are increased<sup>71</sup>. Most recent efforts for enhanced thermoelectrics, however, have been based on reducing the thermal conductivity through nanostructuring. This has opened up promising opportunities to design and fabricate highly efficient thermoelectric materials. The remarkable progress in thermoelectrics is demonstrated by the large number of nanostructures with low thermal conductivities and enhanced thermoelectric efficiencies that have been fabricated<sup>72–79</sup>.

Initial studies indicated that the  $\text{Bi}_2\text{Te}_3/\text{Sb}_2\text{Te}_3$  and  $\text{PbSe}/\text{PbTe}$  quantum-dot superlattices exhibit enhanced thermoelectric properties<sup>72,73</sup>. Their high  $ZT$  values were attributed to the strong reduction in the thermal conductivity. Essentially, studies on superlattices demonstrate that the efficiency of thermoelectrics can be increased by purposely adding interfaces that act as scattering centres for phonons, reducing their mean free paths and therefore the thermal conductivity. In addition to atomic-layer-deposition structures, bulk thermoelectrics made of nanostructured  $\text{AgPbSbTe}$  compounds were shown to provide high  $ZT$  values<sup>74</sup>. In these materials, nanometre-scale quantum dots are produced through a thermal processing technique. In analogy to superlattices, the quantum-dot interfaces strongly scatter phonons, providing a mechanism for reducing the thermal conductivity and increasing the figure of merit  $ZT$ .

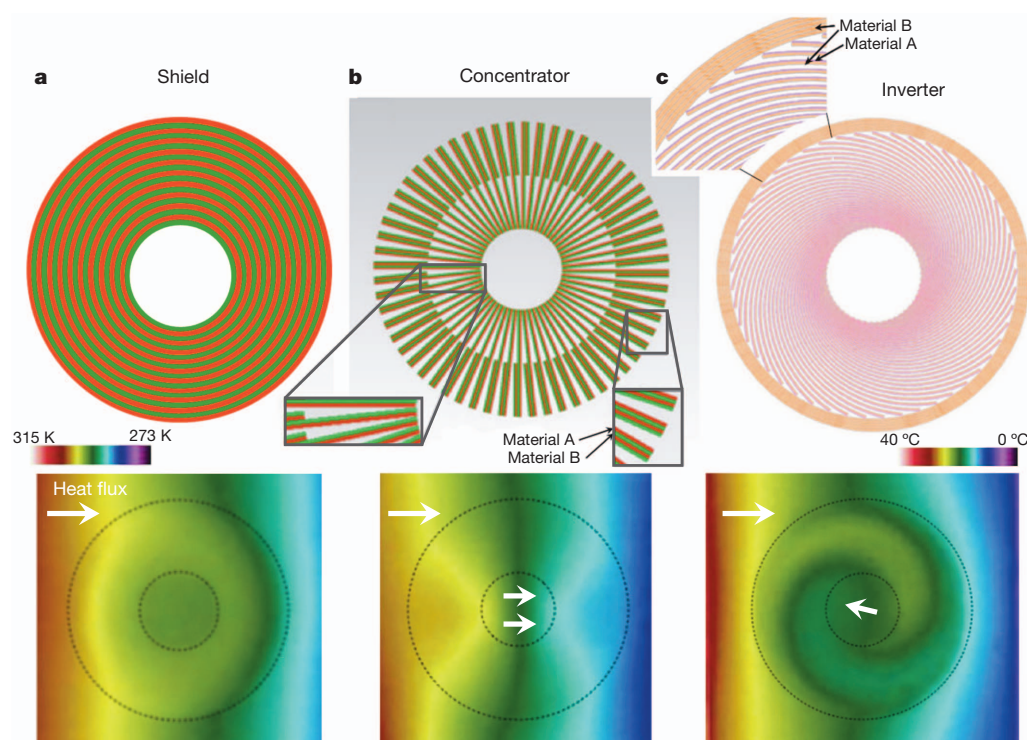


**Figure 6 | Thermal diodes.** **a**, Schematic of a thermal diode concept using two coupled nonlinear one-dimensional lattices. If  $T_L < T_R$ , (where 'L' means left and 'R' means right), vibrations in material B concentrate at low frequencies, but in material A they concentrate at high frequencies. The frequencies do not match, so heat cannot flow efficiently from right to left (reverse direction). If  $T_L > T_R$ , material A expands its vibrations to low frequencies, overlapping with those from material B, and heat is allowed to flow through the

structure (forward direction). Adapted from ref. 63 with permission.

**b**, Schematic and scanning electron microscope image of a boron nitride nanotube thermal diode with asymmetric deposition of amorphous  $\text{C}_9\text{H}_{16}\text{Pt}$ . For asymmetric nanotubes, higher thermal conductances are measured when heat flows from a high-mass region to a low-mass region. Reproduced from fig. 3b of ref. 64 with permission.





**Figure 7 | Thermal metamaterials.** **a**, A thermal shield made of a concentric layered structure of latex rubber and silicone elastomer (top). The measured temperature profile for horizontal heat flux (white arrow): the shell maintains the inner region at constant temperature (bottom). **b**, A thermal concentrator made of azimuthally alternating layers of latex and elastomer (top). The temperature profile, where heat flux is increased by 44% in the inner region (bottom). **c**, A thermal inverter made of a spiral arrangement of copper and polyurethane (top). The temperature profile showing inversion of the heat flux (bottom). Reproduced from figs 2, 3i, 4c and 6a of ref. 69 with permission.

Enhanced thermoelectric efficiencies by reducing heat transport can also be achieved by combining alloys and nanoparticles<sup>75</sup>. Thermal conductivity values below the alloy limit were obtained by epitaxially embedding ErAs nanoparticles in an  $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$  alloy. The basic idea is that the alloy matrix scatters short-wavelength phonons and the ErAs nanoparticles stop mid-to-long-wavelength phonons, reducing the thermal conductivity below the alloy limit and improving thermoelectric efficiency. Another way of reducing phonon transport uses bulk nanocomposites made of BiSbTe alloys with nanosize grains<sup>76</sup>. In this ‘polycrystal’ technique, the material is converted into a nanoparticle dust by ball milling and then hot-pressed such that the nanoparticles merge together, creating nanosize grains. Measurements show<sup>76</sup> that the enhanced thermoelectric properties are the result of a significant reduction in thermal conductivity (the reduction is caused by scattering of phonons at the grain boundaries).

A significant reduction in thermal transport has also been achieved in silicon nanowires<sup>77,78</sup>. Large-area arrays of silicon nanowires grown by electrochemical synthesis show a surprising hundred-fold reduction in thermal conductivity<sup>77</sup>. The main reason for this reduction is the shortening of the phonons’ mean free paths by phonon scattering at the rough nanowire surfaces. The development of theoretical models that can describe such ultralow thermal conductivities is currently an intense research area ranging from atomistic calculations to Boltzmann transport models.

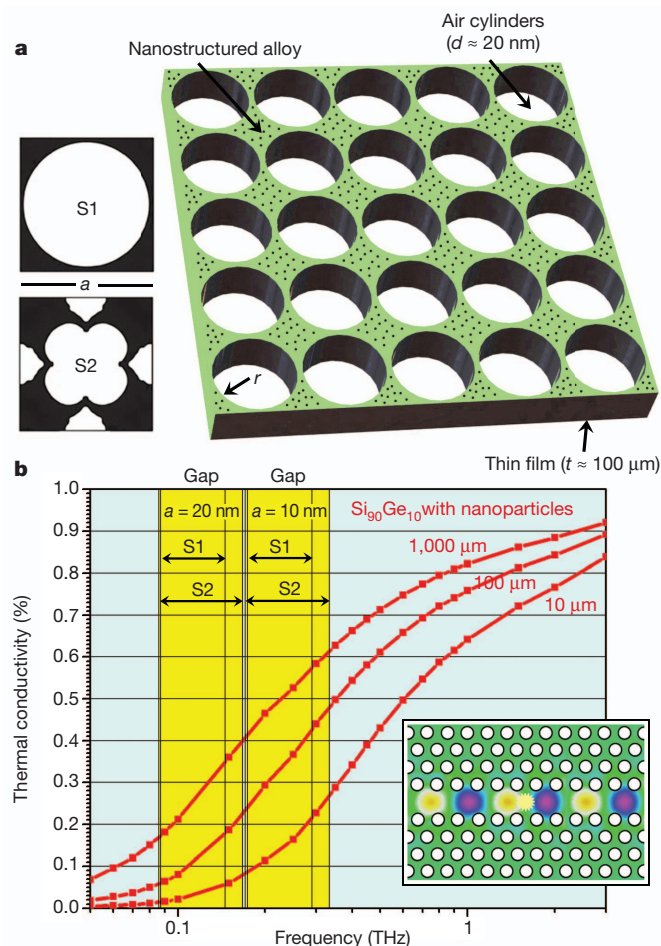
More recently, the combination of various scattering processes has been employed to reduce heat transport greatly in PbTe-based thermoelectric materials<sup>79</sup>. First, atomic-scale substitutions (2 mol.% doping with Na) are introduced in a bulk PbTe sample to generate a figure of merit  $ZT \approx 1.1$  at 775 K. By including endotaxial SrTe nanoparticles with 2–10 nm sizes,  $ZT$  is increased to about 1.7 at 800 K. Then, by transforming the sample to a polycrystal material with micrometre grain sizes, a  $ZT$  value of about 2.2 is achieved at 915 K (ref. 79). The nanocomposite provides a high  $ZT$  owing to the different length scales in the structure, which scatter phonons with different wavelengths. Essentially, atomic-scale substitutions block short-wavelength phonons, nanoparticles block mid-wavelength phonons and micrograins block large-wavelength phonons. These combined processes create an ultralow thermal conductivity and a highly efficient thermoelectric. The Na-doped PbTe–SrTe matrix was synthesized by melting and quenching, while the polycrystal was produced by powder processing and spark plasma sintering.

The idea of reducing heat transport through nanostructuring has been the basis for many recent advances in thermoelectric materials. The search for an easy-to-craft nanostructure with low thermal conductivity is now one of the main challenges in thermoelectrics. Equally importantly, nanostructured materials show us that interface scattering is a very effective mechanism of stopping the flow of phonons in the high-frequency (terahertz) part of the phononic spectrum<sup>72–80</sup>.

### Thermocrystals and coherent heat effects

An entirely new way of controlling heat transport, by which heat flow can be managed like sound waves<sup>81</sup>, has recently been developed using ‘thermocrystals’, which are periodic structures made of alloys containing nanoparticles (Fig. 8a). The basic idea of thermocrystals is to manipulate the heat frequency spectrum such that low-frequency phonons carry a considerable part of the heat. In this manner, heat-carrying phonons are subject to coherent reflection and transmission at interfaces, and thus many applications designed for sound management can be applied to heat flow.

In thermocrystals<sup>81</sup>, high-frequency phonons are blocked by including alloy atoms and nanoparticles ( $d \approx 1$  nm). The alloy atoms scatter high-frequency phonons, while the nanoparticles block another set of high-frequency phonons<sup>82–84</sup>. In contrast to thermoelectrics<sup>82</sup>, very high-frequency phonons are blocked in thermocrystals<sup>81</sup>. Because high-frequency phonons are severely restricted from carrying heat, the proportion of heat carried by low-frequency phonons increases (Fig. 8b). To concentrate heat around a specific frequency range—with the purpose of matching the PBG—the contribution of very low-frequency phonons is also reduced by considering a thin-film material with rough surfaces. As a result of blocking the highest and lowest frequencies, most heat is thus concentrated to a relatively narrow low-frequency band. Specifically, for a  $\text{Si}_{90}\text{Ge}_{10}$  alloy with a 10% filling fraction of  $d \approx 1$  nm Ge nanoparticles, the frequency band is  $0.1 \text{ THz} < f < 2 \text{ THz}$ , with up to 40% of the heat restricted to a narrow ‘hypersonic’ 100–300-GHz range (Fig. 8b). Taking advantage of this reduced heat spectrum, phononic crystals<sup>12,85</sup> with periodicities  $a \approx 10$ –20 nm are designed such that their PBGs match the frequencies of this narrow hypersonic range. These crystals are created by introducing air holes in the SiGe alloy thin film containing nanoparticles (Fig. 8a). In these ‘thermocrystals’ up to 23% of heat is carried by phonons with frequencies within the PBG.



**Figure 8 | Thermocrystals.** **a**, Schematic of a thermocrystal made of air cylinders on a nanostructured alloy film<sup>81</sup>. **b**, Red shows the cumulative thermal conductivity as a function of frequency for  $\text{Si}_{90}\text{Ge}_{10}$  nanoparticle films with different thicknesses. Yellow shows the frequency bandgaps for films patterned with a square arrangement of cylinders (S1) and with the design (S2). The inset shows that among the opportunities provided by thermocrystals is the possibility of guiding heat by removing a rod of cylinders. The colour scheme shows the displacement for a heat-carrying low-frequency phonon propagating within the waveguide.

This matching of the heat frequency spectrum and the PBG demonstrates that thermal phonons can be managed through coherent interference effects. In the past, phononic crystals have been shown to manage sound waves, but these are the first phononic crystals designed to control heat-carrying phonons. Thermocrystals lay the foundation for various applications such as heat waveguides, thermal lattices, heat imaging, thermo-optics, thermal diodes and thermal cloaking.

## Conclusions and perspectives

Newly developed phononic crystals and metamaterials are both able to control phonon transport successfully at low and high frequencies in the phononic spectrum, ranging from sound to heat transfer. There is, however, an important difference between phonon management based on phononic crystals and that based on metamaterials. Whereas in phononic crystals the structure periodicity is generally of the same order as the wavelengths of the phonons to be controlled, in metamaterials the structure length scale can be smaller. This means that if we want to control large wavelengths in the phononic spectrum (Fig. 1), such as earthquake or tsunami waves, it would be more appropriate to use metamaterials, because the required periodicities for phononic crystals would be exceedingly large. On the other hand, if we want to manipulate the short-wavelength limit, that is, heat flow, then both thermocrystals

and thermal metamaterials may be appropriate choices depending on the structure length scale and intended applications, because heat flow can be controlled by using millimetre-scale metamaterials or nanometre-scale thermocrystals.

I have also shown that heat flow can be blocked to ultralow levels using a large variety of nanostructured materials. This large reduction of the thermal conductivity is a consequence of diffuse interface scattering. When the nanostructured material is periodic, however, heat flow can be affected by two different physical mechanisms: diffuse interface scattering or coherent wave interference. For a periodic nanostructure to control heat coherently—and act as a phononic crystal—the wavelengths of heat-carrying phonons must be similar to the structure periodicity,  $\lambda \approx 2a$  (Bragg reflection). Given that most heat phonon wavelengths in standard semiconductors at room temperature are relatively short ( $\lambda \approx 1$ –10 nm in silicon), the Bragg condition is essentially not satisfied in typical periodic nanostructures, and this prevents heat manipulation by PBGs. By increasing heat phonon wavelengths using thermocrystals, however, coherent wave-interference effects (and heat wave transport) can be achieved in nanostructures that are periodic.

This Review has shown that the development of new ideas for phonon management—combined with the ability to design and fabricate composite materials from the macroscale to the nanometre scale—has fuelled recent progress in sonic and thermal diodes, acoustic and thermal metamaterials, optomechanical crystals, hypersonic phononic crystals, thermoelectrics and thermocrystals. These advances have greatly increased our ability to manage the phononic spectrum at all relevant frequencies: sound, ultrasound, hypersound and heat. The emerging field of phonon management has great potential for innovations in materials and devices that can precisely manipulate sound and heat. Our ability to control electrons and photons has driven major technological revolutions in past decades; perhaps from our new ability to control phonons precisely we may expect analogously surprising and exciting consequences.

Received 19 April; accepted 28 August 2013.

- Kushwaha, M. S., Halevi, P., Dobrzynski, L. & Djafari-Rouhani, B. Acoustic band structure of periodic elastic composites. *Phys. Rev. Lett.* **71**, 2022–2025 (1993).
- Sigalas, M. M. & Economou, E. N. Band structure of elastic waves in two dimensional systems. *Solid State Commun.* **86**, 141–143 (1993).
- Martínez-Sala, R. *et al.* Sound attenuation by sculpture. *Nature* **378**, 241 (1995).
- Sanchez-Perez, J. V. *et al.* Sound attenuation by a two-dimensional array of rigid cylinders. *Phys. Rev. Lett.* **80**, 5325–5328 (1998).
- Montero de Espinosa, F. R., Jimenez, E. & Torres, M. Ultrasonic band gap in a periodic two-dimensional composite. *Phys. Rev. Lett.* **80**, 1208–1211 (1998).
- Liu, Z. Y. *et al.* Locally resonant sonic materials. *Science* **289**, 1734–1736 (2000).
- Vasseur, J. O. *et al.* Experimental and theoretical evidence for the existence of absolute acoustic band gaps in two-dimensional solid phononic crystals. *Phys. Rev. Lett.* **86**, 3012–3015 (2001).
- Gorishnyy, T., Ullal, C. K., Maldovan, M., Fytas, G. & Thomas, E. L. Hypersonic phononic crystals. *Phys. Rev. Lett.* **94**, 115501 (2005).
- This paper describes the experimental realization of small-scale phononic crystals that control high-frequency hypersonic phonons.**
- Cheng, W., Wang, J., Jonas, U., Fytas, G. & Stefanou, N. Observation and tuning of hypersonic bandgaps in colloidal crystals. *Nature Mater.* **5**, 830–836 (2006).
- Thomas, E. L., Gorishnyy, T. & Maldovan, M. Phononics: colloidal crystals go hypersonic. *Nature Mater.* **5**, 773–774 (2006).
- Yu, J.-K., Mitrovic, S., Tham, D., Varghese, J. & Heath, J. R. Reduction of thermal conductivity in phononic nanomesh structure. *Nature Nanotechnol.* **5**, 718–721 (2010).
- Maldovan, M. & Thomas, E. L. Simultaneous localization of phonons and photons in two-dimensional periodic structures. *Appl. Phys. Lett.* **88**, 251907 (2006).
- Liang, B., Yuan, B. & Cheng, J. C. Acoustic diode: rectification of acoustic energy flux in one-dimensional systems. *Phys. Rev. Lett.* **103**, 104301 (2009).
- Liang, B., Guo, X. S., Tu, J., Zhang, D. & Chen, J. C. An acoustic rectifier. *Nature Mater.* **9**, 989–992 (2010).
- Li, B. Acoustics: now you hear me, now you don't. *Nature Mater.* **9**, 962–963 (2010).
- Li, X.-F. *et al.* Tunable unidirectional sound propagation through a sonic-crystal-based acoustic diode. *Phys. Rev. Lett.* **106**, 084301 (2011).
- This paper describes the experimental realization of an acoustic diode by breaking spatial inversion symmetry in phononic crystals.**
- Boechler, N., Theocaris, G. & Daraio, C. Bifurcation-based acoustic switching and rectification. *Nature Mater.* **10**, 665–668 (2011).



18. Pendry, J. B., Schurig, D. & Smith, D. R. Controlling electromagnetic fields. *Science* **312**, 1780–1782 (2006).
19. Leonhardt, U. Optical conformal mapping. *Science* **312**, 1777–1780 (2006).
20. Milton, G. W., Briane, M. & Willis, J. R. On cloaking for elasticity and physical equations with a transformation invariant form. *New J. Phys.* **8**, 248 (2006).
21. Cummer, S. A. & Schurig, D. One path to acoustic cloaking. *New J. Phys.* **9**, 45 (2007).
22. Chen, H. & Chan, C. T. Acoustic cloaking in three dimensions using acoustic metamaterials. *Appl. Phys. Lett.* **91**, 183518 (2007).
23. Cummer, S. A. *et al.* Scattering theory derivation of a 3D acoustic cloaking shell. *Phys. Rev. Lett.* **100**, 024301 (2008).
24. Chen, H. & Chan, C. T. Acoustic cloaking and transformation acoustics. *J. Phys. D* **43**, 113001 (2010).
25. Torrent, D. & Dehesa-Sanchez, J. Acoustic cloaking in two-dimensions: a feasible approach. *New J. Phys.* **10**, 063015 (2008).
26. Cheng, Y., Yang, F., Xu, J. Y. & Liu, X. J. A multilayer structured acoustic cloak with homogeneous isotropic materials. *Appl. Phys. Lett.* **92**, 151913 (2008).
27. Zhang, S., Cia, X. & Fang, N. Broadband acoustic cloak for ultrasound waves. *Phys. Rev. Lett.* **106**, 024301 (2011).
- This paper describes the experimental realization of acoustic cloaking shells for ultrasound waves using purpose-designed metamaterials.**
28. Chan, C. T. Invisibility cloak for ultrasonic waves. *Physics* **4**, 2 (2011).
29. Farhat, M., Enoch, S., Guenneau, S. & Movchan, A. B. Broadband cylindrical acoustic cloak for linear surface waves in a fluid. *Phys. Rev. Lett.* **101**, 134501 (2008).
30. Popa, B. I., Zigoneanu, L. & Cummer, S. A. Experimental acoustic ground cloak in air. *Phys. Rev. Lett.* **106**, 253901 (2011).
31. Stenger, N., Wilhelm, M. & Wegener, M. Experiments on elastic cloaking in thin plates. *Phys. Rev. Lett.* **108**, 014301 (2012).
32. Farhat, M., Guenneau, S. & Enoch, S. Ultrabroadband elastic cloaking in thin plates. *Phys. Rev. Lett.* **103**, 024301 (2009).
33. Brun, M., Guenneau, S. & Movchan, A. B. Achieving control of in-plane elastic waves. *Appl. Phys. Lett.* **94**, 061903 (2009).
34. Trigo, M., Bruchhausen, A., Fainstein, A., Jusserand, B. & Thieryy-Mieg, V. Confinement of acoustical vibrations in semiconductor planar phonon cavity. *Phys. Rev. Lett.* **89**, 227402 (2002).
35. Worlock, J. M. & Roukes, M. L. Son et lumière. *Nature* **421**, 802–803 (2003).
36. Yablonovitch, E. Inhibited spontaneous emission in solid-state physics and electronics. *Phys. Rev. Lett.* **58**, 2059–2062 (1987).
37. John, S. Strong localization of photons in certain disordered dielectric superlattices. *Phys. Rev. Lett.* **58**, 2486–2489 (1987).
38. Joannopoulos, J. D., Villeneuve, P. R. & Fan, S. Photonic crystals: putting a new twist on light. *Nature* **386**, 143–149 (1997).
39. Maldovan, M. & Thomas, E. L. *Periodic Structures and Interference Lithography: for Photonics, Phononics and Mechanics* (Wiley, 2008).
40. Eichenfield, M., Chan, J., Camacho, R. M., Vahala, K. J. & Painter, O. Optomechanical crystals. *Nature* **462**, 78–82 (2009).
- This paper describes the experimental demonstration of phonon-photon coupling in planar 'optomechanical' crystals.**
41. Psarobas, I. E. *et al.* Enhanced acousto-optic interactions in a one-dimensional phoxonic cavity. *Phys. Rev. B* **82**, 174303 (2010).
42. Fainstein, A., Lanzillotti-Kimura, N. D., Jusserand, B. & Perrin, B. Strong optical-mechanical coupling in a vertical GaAs/AlAs microcavity for subterahertz phonons and near-infrared light. *Phys. Rev. Lett.* **110**, 037403 (2013).
43. Sadat-Saleh, S., Benchabane, S., Baida, F. I., Bernal, M. P. & Laude, V. Tailoring simultaneous photonic and phononic band gaps. *J. Appl. Phys.* **106**, 074912 (2009).
44. Mohammadi, S., Eftekhari, A. A., Khelif, A. & Adibi, A. Simultaneous two-dimensional phononic and photonic band gaps in opto-mechanical crystal slabs. *Opt. Express* **18**, 9164–9172 (2010).
45. Pennec, Y. *et al.* Simultaneous existence of phononic and photonic bandgaps in periodic crystal slabs. *Opt. Express* **18**, 14301–14310 (2010).
46. Safavi-Naeini, A. H. & Painter, O. Design of optomechanical cavities and waveguides on a simultaneous bandgap phononic-photonic crystal slab. *Opt. Express* **18**, 14926–14943 (2010).
47. Safavi-Naeini, A. H., Mayer Alegre, T. P., Winger, M. & Painter, O. Optomechanics in an ultrahigh-Q two-dimensional photonic crystal cavity. *Appl. Phys. Lett.* **97**, 181106 (2010).
48. Gavartin, E. *et al.* Optomechanical coupling in a two-dimensional photonic crystal defect cavity. *Phys. Rev. Lett.* **106**, 203902 (2011).
49. Chan, J. *et al.* Laser cooling of a nanomechanical oscillator into its quantum ground state. *Nature* **478**, 89–92 (2011).
50. Safavi-Naeini, A. H. *et al.* Observation of quantum motion of a nanomechanical resonator. *Phys. Rev. Lett.* **108**, 033602 (2012).
51. Safavi-Naeini, A. H. *et al.* Electromagnetically induced transparency and slow light with optomechanics. *Nature* **472**, 69–73 (2011).
52. Heinrich, G., Ludwig, M., Qian, J., Kubala, B. & Marquardt, F. Collective dynamics of optomechanical arrays. *Phys. Rev. Lett.* **107**, 043603 (2011).
53. Grudinin, I. S., Lee, H., Painter, O. & Vahala, K. J. Phonon laser action in a tunable two-level system. *Phys. Rev. Lett.* **104**, 083901 (2010).
54. Akimov, A. V. *et al.* Hypersonic modulation of light in three-dimensional photonic and phononic band gap materials. *Phys. Rev. Lett.* **101**, 033902 (2008).
55. Papanikolaou, N., Psarobas, I. E. & Stefanou, N. Absolute spectral gaps for infrared light and hypersound in three-dimensional metallo-dielectric phoxonic crystals. *Appl. Phys. Lett.* **96**, 231917 (2010).
56. Russell, P. S. J., Marin, E., Diez, A., Guenneau, S. & Movchan, A. B. Sonic band gaps in PCF preforms: enhancing the interaction of sound and light. *Opt. Express* **11**, 2555–2560 (2003).
57. Laude, V. *et al.* Phononic bandgap guidance of acoustic modes in photonic crystal fibers. *Phys. Rev. B* **71**, 045107 (2005).
58. Dainese, P. *et al.* Stimulated Brillouin scattering from multi-GHz-guided acoustic phonons in nanostructured photonic crystal fibres. *Nature Phys.* **2**, 388–392 (2006).
59. Kang, M. S., Nazarkin, A., Brenn, A. & Russell, P. S. J. Tightly trapped acoustic phonons in photonic crystal fibres as highly nonlinear artificial Raman oscillators. *Nature Phys.* **5**, 276–280 (2009).
60. Li, B. W., Wang, L. & Casati, G. Thermal diode: rectification of heat flux. *Phys. Rev. Lett.* **93**, 184301 (2004).
61. Terraneo, M., Peyrard, M. & Casati, G. Controlling the energy flow in nonlinear lattices: a model for a thermal rectifier. *Phys. Rev. Lett.* **88**, 094302 (2002).
62. Wang, L. & Li, B. Thermal logic gates: computation with phonons. *Phys. Rev. Lett.* **99**, 177208 (2007).
63. Wang, L. & Li, B. Phononics get hot. *Phys. World* **21**, 27–29 (2008).
64. Chang, C. W., Okawa, D., Majumdar, A. & Zettl, A. Solid-state thermal rectifier. *Science* **314**, 1121–1124 (2006).
- This paper describes an experimental solid-state thermal diode formed by a non-uniform mass distribution in nanotubes.**
65. Yang, N., Li, N., Wang, L. & Li, B. Thermal rectification and negative differential thermal resistance in lattices with mass gradient. *Phys. Rev. B* **76**, 020301 (2007).
66. Fan, C. Z., Gao, Y. & Huang, J. P. Shaped graded materials with an apparent negative thermal conductivity. *Appl. Phys. Lett.* **92**, 251907 (2008).
67. Chen, T., Weng, C. N. & Chen, J. S. Cloak for curvilinearly anisotropic media in conduction. *Appl. Phys. Lett.* **93**, 114103 (2008).
68. Guenneau, S., Amra, C. & Veynante, D. Transformation thermodynamics: cloaking and concentrating heat flux. *Opt. Express* **20**, 8207–8218 (2012).
69. Narayana, S. & Sato, Y. Heat flux manipulation by engineered thermal materials. *Phys. Rev. Lett.* **108**, 214303 (2012).
- This paper describes the manipulation of heat conduction using newly developed thermal metamaterials.**
70. Schittny, R., Kadic, M., Guenneau, S. & Wegener, M. Experiments on transformation thermodynamics: molding the flow of heat. *Phys. Rev. Lett.* **110**, 195901 (2013).
71. Hicks, L. D. & Dresselhaus, M. S. Effect of quantum-well structures on the thermoelectric figure of merit. *Phys. Rev. B* **47**, 12727–12731 (1993).
72. Venkatasubramanian, R., Siivola, E., Colpitts, T. & O'Quinn, B. Thin-film thermoelectric devices with high-room temperature figures of merit. *Nature* **413**, 597–602 (2001).
73. Harman, T. C., Taylor, P. J., Walsh, M. P. & LaForge, B. E. Quantum dot superlattice thermoelectric materials and devices. *Science* **297**, 2229–2232 (2002).
74. Hsu, K. F. *et al.* Cubic AgPb<sub>m</sub>SbTe<sub>2+m</sub>: bulk thermoelectric materials with high figure of merit. *Science* **303**, 818–821 (2004).
75. Kim, W. *et al.* Thermal conductivity reduction and thermoelectric figure of merit increase by embedding nanoparticles in crystalline semiconductors. *Phys. Rev. Lett.* **96**, 045901 (2006).
76. Poudel, B. *et al.* High-thermoelectric performance of nanostructured bismuth antimony telluride bulk alloys. *Science* **320**, 634–638 (2008).
77. Hochbaum, A. I. *et al.* Enhanced thermoelectric performance of rough silicon nanowires. *Nature* **451**, 163–167 (2008).
78. Boukai, A. I. *et al.* Silicon nanowires as efficient thermoelectric materials. *Nature* **451**, 168–171 (2008).
79. Biswas, K. *et al.* High-performance bulk thermoelectrics with all-scale hierarchical structures. *Nature* **489**, 414–418 (2012).
- This paper describes a highly efficient thermoelectric material obtained through the scattering of a wide range of phonons with different wavelengths.**
80. Chiriac, C. *et al.* Ultra low thermal conductivity in disordered WSe<sub>2</sub> crystals. *Science* **315**, 351–353 (2007).
81. Maldovan, M. Narrow low-frequency spectrum and heat management by thermocrystals. *Phys. Rev. Lett.* **110**, 025902 (2013).
82. Mingo, N., Hauser, D., Kobayashi, N. P., Plissonnier, M. & Shakouri, A. Nanoparticle in alloy approach to efficient thermoelectrics: silicides in SiGe. *Nano Lett.* **9**, 711–715 (2009).
83. Garg, J., Bonini, N., Kozinsky, B. & Marzari, N. Role of disorder and anharmonicity in the thermal conductivity of silicon germanium alloys: a first principle study. *Phys. Rev. Lett.* **106**, 045901 (2011).
84. Kundu, A., Mingo, N., Broido, D. A. & Stewart, D. A. Role of light and heavy embedded nanoparticles on the thermal conductivity of SiGe alloys. *Phys. Rev. B* **84**, 125426 (2011).
85. Bilal, O. R. & Hussein, M. I. Ultrawide phononic band gap for combined in-plane and out-of-plane waves. *Phys. Rev. E* **84**, 065701 (2011).

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The author declares no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.M. (maldovan@gatech.edu).

# Dedifferentiation of committed epithelial cells into stem cells *in vivo*

Purushothama Rao Tata<sup>1,2,3,4</sup>, Hongmei Mou<sup>1,2,3,4</sup>, Ana Pardo-Saganta<sup>1,2,3,4</sup>, Rui Zhao<sup>1,2,3,4</sup>, Mythili Prabhu<sup>1,2,3,4</sup>, Brandon M. Law<sup>1,2,3,4</sup>, Vladimir Vinarsky<sup>1,2,3,4</sup>, Josalyn L. Cho<sup>3,5</sup>, Sylvie Breton<sup>6</sup>, Amar Sahay<sup>1,4,7</sup>, Benjamin D. Medoff<sup>3,5</sup> & Jayaraj Rajagopal<sup>1,2,3,4</sup>

Cellular plasticity contributes to the regenerative capacity of plants, invertebrates, teleost fishes and amphibians. In vertebrates, differentiated cells are known to revert into replicating progenitors, but these cells do not persist as stable stem cells. Here we present evidence that differentiated airway epithelial cells can revert into stable and functional stem cells *in vivo*. After the ablation of airway stem cells, we observed a surprising increase in the proliferation of committed secretory cells. Subsequent lineage tracing demonstrated that the luminal secretory cells had dedifferentiated into basal stem cells. Dedifferentiated cells were morphologically indistinguishable from stem cells and they functioned as well as their endogenous counterparts in repairing epithelial injury. Single secretory cells clonally dedifferentiated into multipotent stem cells when they were cultured *ex vivo* without basal stem cells. By contrast, direct contact with a single basal stem cell was sufficient to prevent secretory cell dedifferentiation. In analogy to classical descriptions of amphibian nuclear reprogramming, the propensity of committed cells to dedifferentiate is inversely correlated to their state of maturity. This capacity of committed cells to dedifferentiate into stem cells may have a more general role in the regeneration of many tissues and in multiple disease states, notably cancer.

The term dedifferentiation was first coined to describe the process in which cells of the urodele retinal pigment epithelium lose their differentiated properties to replace extirpated lens cells<sup>1</sup>. Although not formally demonstrated, the term was used to suggest that differentiated epithelial cells reverted to a previous developmental stage before their subsequent differentiation into an alternative cell fate. Dedifferentiation has since been explored in plants, invertebrates, teleost fishes and amphibians<sup>2–17</sup>. In vertebrates, quiescent differentiated cells can revert into replicating progenitor cells<sup>5–7,11,12,14</sup> to replace lost cells, but these progenitor cells do not persist as stable stem cells<sup>11</sup>. Indeed, in murine hair follicle regeneration, the immediate differentiated progeny of epithelial stem cells are already resistant to dedifferentiation<sup>17</sup>. Conversely, the undifferentiated secretory progenitors of the intestine that are the immediate progeny of intestinal stem cells are able to dedifferentiate into stem cells after injury<sup>13</sup>, mimicking the capacity for dedifferentiation of the immediate progeny of *Drosophila* germline stem cells<sup>3,15,16</sup>. Recently, using stringent lineage-tracing strategies<sup>18</sup>, airway epithelial cells have been shown to be more plastic than recognized previously. In a separate study, differentiated secretory cells have been shown to give rise to very rare cells ( $0.34 \pm 0.09\%$ ) that express basal cell markers after severe injury, but the properties of these rare basal-like cells were not studied and their functional capacity was not assessed<sup>19</sup>. Here, we specifically sought to determine whether stably committed luminal cells could dedifferentiate into functional stem cells.

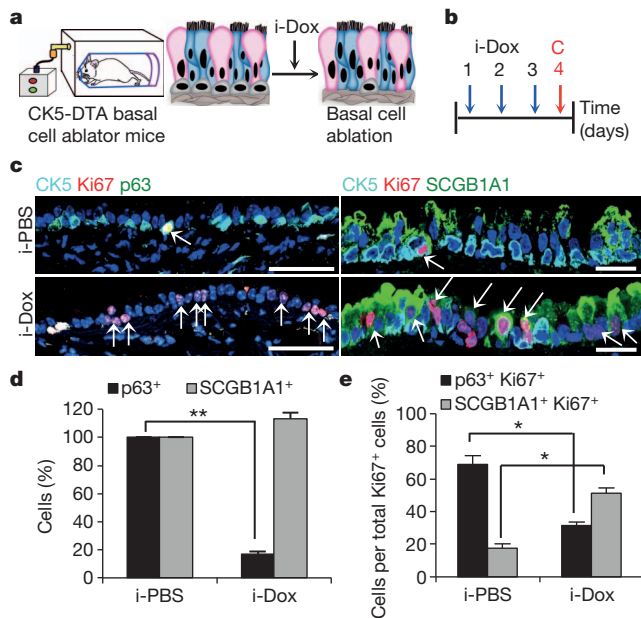
## Secretory cells replicate after stem cell ablation

Airway basal stem cells have been shown to self-renew and differentiate into multiple airway epithelial cell types using genetic lineage tracing<sup>20,21</sup>. Secretory cells are differentiated luminal cells that have both secretory and detoxifying functions. Secretory cells can also

further differentiate into ciliated cells<sup>19</sup>. To test whether secretory cells can dedifferentiate into stem cells, we ablated basal stem cells of the airway epithelium and simultaneously lineage traced the secretory cells of the same mouse (Extended Data Fig. 1). To ablate the airway basal stem cells, we generated a *CK5-rtTA/tet(O)DTA*<sup>22,23</sup> (hereafter referred to as CK5-DTA) mouse, which expresses the active subunit of the diphtheria toxin (DTA) in CK5 (also known as KRT5)<sup>+</sup> airway basal stem cells following doxycycline administration. CK5 expression is, however, not restricted to the basal stem cells of the airway epithelium and is expressed in many other epithelial tissues<sup>20,22</sup>. Therefore, the ablation of CK5-expressing cells using systemic doxycycline administration through drinking water or by intraperitoneal injection is lethal in the adult mouse (data not shown). To circumvent this problem, we designed a method for the inhalational delivery of doxycycline<sup>24</sup> as a means of inducing DTA transgene expression exclusively in the airway epithelium. CK5-DTA mice were exposed to either inhaled PBS (i-PBS) or inhaled doxycycline (i-Dox) for two or three consecutive days (Fig. 1a, b). Tracheas were isolated 24 h after the final dose of i-Dox. Immunofluorescence analysis using the basal stem cell markers p63 (also known as TP63), CK5, NGFR and T1 $\alpha$  (also known as PDPN) demonstrated a specific dose-dependent ablation of basal cells and a preservation of normal secretory cell numbers (Fig. 1c, d and Extended Data Fig. 2a–c). Approximately 80% ( $n = 3$ ) of the airway basal cells were ablated after three doses of inhaled doxycycline (an  $81.0 \pm 3.4\%$  decrease in the number of CK5<sup>+</sup> cells and a  $78.8 \pm 2.4\%$  decrease in the number of p63<sup>+</sup> cells) (Extended Data Fig. 2d). Immunofluorescence staining of the proliferation marker Ki67 in combination with cell-type-specific markers demonstrated that residual CK5<sup>+</sup> and p63<sup>+</sup> basal cells do proliferate, but there are actually fewer numbers of proliferating basal cells relative

<sup>1</sup>Center for Regenerative Medicine, Massachusetts General Hospital, 185 Cambridge Street, Boston, Massachusetts 02114, USA. <sup>2</sup>Departments of Pediatrics, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. <sup>3</sup>Department of Internal Medicine, Pulmonary and Critical Care Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. <sup>4</sup>Harvard Stem Cell Institute, Cambridge, Massachusetts 02138, USA. <sup>5</sup>Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA. <sup>6</sup>Center for Systems Biology, Program in Membrane Biology and Nephrology Division, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02214, USA. <sup>7</sup>Department of Psychiatry, Harvard Medical School, Boston, Massachusetts 02215, USA.



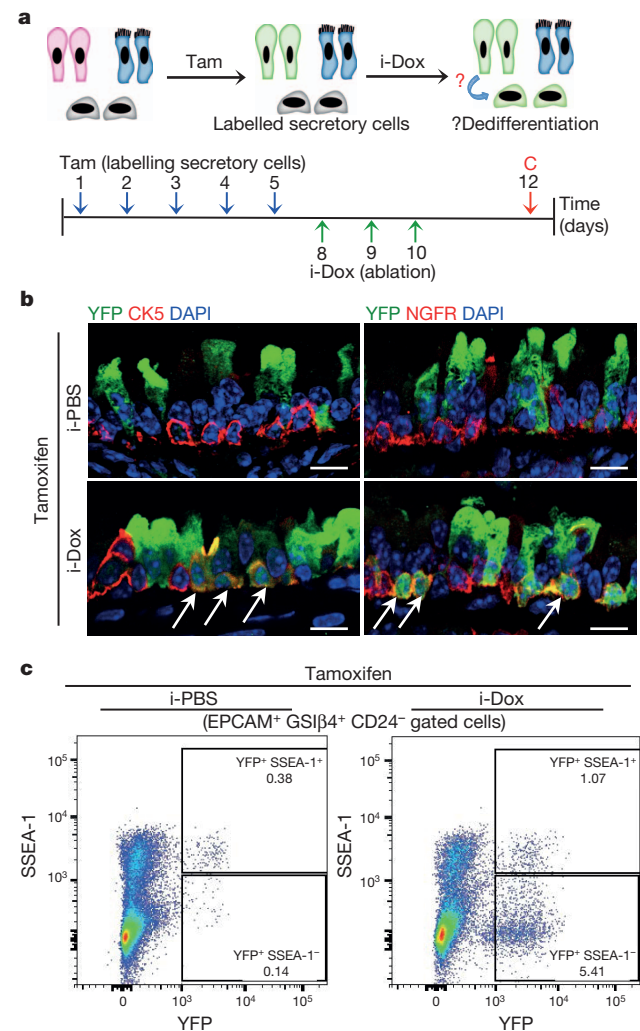


**Figure 1 | Secretory cells proliferate after basal cell ablation.** **a**, Schematic representation of the ablation of CK5-expressing basal cells of the trachea. Secretory, ciliated and basal stem cells are shown in pink, blue and grey, respectively. **b**, Schematic of the timeline of i-Dox or i-PBS administration and tissue collection (C). **c**, Immunostaining for basal (p63 (green) and CK5 (cyan)) and secretory cells (SCGB1A1 (green)) in combination with Ki67 (red) on either i-PBS (top)- or i-Dox (bottom)-treated mice ( $n = 6$ ). White arrows, Ki67<sup>+</sup> cells. **d**, Quantification of the percentage of p63<sup>+</sup> and SCGB1A1<sup>+</sup> cells per total number of DAPI (4',6-diamidino-2-phenylindole)-stained epithelial cells in i-PBS or i-Dox groups.  $n = 3$ . **e**, Percentage of p63<sup>+</sup> Ki67<sup>+</sup> and SCGB1A1<sup>+</sup> Ki67<sup>+</sup> cells relative to total Ki67<sup>+</sup> cells in i-PBS- and i-Dox ( $n = 3$ )-treated CK5-DTA mice. Nuclei, DAPI (blue). \* $P < 0.05$ , \*\* $P < 0.01$  (two tailed and paired  $t$ -test).  $n = 3$  (three mice per condition). Error bars, average  $\pm$  s.e.m. Scale bars, 20  $\mu$ m.

to the total population of replicating cells after ablation (Fig. 1c, e). Notably, there was a twofold increase in the numbers of replicating secretory cells (SCGB1A1<sup>+</sup> Ki67<sup>+</sup>) in i-Dox-treated animals ( $51.29 \pm 3.02\%$ ) as compared to i-PBS-treated animals ( $17.7 \pm 2.68\%$ ) (Fig. 1c, e and Extended Data Fig. 3a). Consistent with the increased proliferation of differentiated secretory cells, Ki67 staining was specifically increased in the CK8 (also known as KRT8)<sup>+</sup> suprabasal layer of the airway epithelium (Extended Data Fig. 3b). Thus, secretory cells are the predominant cells that replicate after stem cell ablation. Interestingly, occasional cells expressed both the basal cell marker CK5 and the secretory cell marker SCGB1A1 in i-Dox-treated mice (Extended Data Fig. 3c).

### Secretory cells dedifferentiate in vivo

To lineage-label secretory cells before stem cell ablation, we generated quadruple transgenic mice: *Scgb1a1-creER/LSL-YFP::CK5-rtTA-tet(O)DTA* (hereafter referred to as SCGB1A1-YFP/CK5-DTA mice). Administration of tamoxifen to induce the CreER-mediated expression of the yellow fluorescent protein (YFP) label in secretory cells was followed by three doses of i-Dox to induce basal cell ablation (Fig. 2a). Lineage-labelled YFP<sup>+</sup> secretory cells demonstrated increased rates of proliferation in i-Dox-treated animals as compared to i-PBS treated controls (Extended Data Fig. 3d, e). We identified YFP<sup>+</sup> secretory-cell-derived cells that were morphologically indistinguishable from basal stem cells (Fig. 2b). In addition, we found that a subset of lineage-labelled cells expressed a suite of basal cell markers including CK5, NGFR, p63 and T1 $\alpha$  (Fig. 2b and Extended Data Fig. 3f). Quantification revealed that  $7.9 \pm 2.08\%$  of basal cells (585 CK5<sup>+</sup> YFP<sup>+</sup> cells out of 7,320 total CK5<sup>+</sup> cells in i-Dox-treated animals,  $n = 6$  mice) expressed a YFP lineage label,



**Figure 2 | Luminal secretory cells dedifferentiate into basal stem cells after stem-cell ablation.** **a**, Schematic representation of tamoxifen (Tam) and i-Dox administration to SCGB1A1-YFP/CK5-DTA mice followed by tissue collection (C). **b**, Immunostaining for CK5 (red; left) and NGFR (red; right) in combination with YFP (green) in i-PBS (top)- or i-Dox (bottom)-treated mice ( $n = 3$ ). White arrows indicate double-positive cells. **c**, Flow-cytometric analysis of dedifferentiated cells. EPCAM<sup>+</sup> GSI $\beta$ 4<sup>+</sup> CD24<sup>-</sup> cells were analysed for expression of SSEA-1 and YFP in either i-Dox- or i-PBS-treated mice.  $n = 3$  (at least two mice each per condition). Scale bars, 20  $\mu$ m.

demonstrating that dedifferentiated basal-like cells comprised a substantial fraction of the total stem cell pool. Dedifferentiated cells did not appear in PBS-treated controls (3 CK5<sup>+</sup> YFP<sup>+</sup> cells out of 7,558 total CK5<sup>+</sup> cells counted ( $0.041 \pm 0.028\%$ ;  $n = 6$  mice)). Consistently, when the entire basal cell population is purified by flow cytometry, the YFP-lineage-labelled basal-like cells have lost the secretory cell surface marker SSEA-1 (Fig. 2c). Thus, dedifferentiating cells lose markers of secretory cell differentiation as they acquire markers of basal stem cells.

### Secretory cells dedifferentiate ex vivo

Because our *in vivo* experiments demonstrated that secretory cells are stimulated to dedifferentiate by the ablation of basal stem cells, we wondered whether secretory cell dedifferentiation could be induced *ex vivo* when secretory cells were cultured in the absence of basal stem cells. We reasoned that such an assay would provide a platform for further determining whether the dedifferentiation process is actively suppressed by the presence of co-cultured basal stem cells. To assess this possibility, we isolated and sorted unlabelled basal stem cells and YFP<sup>+</sup> secretory cells from SCGB1A1-YFP mice after tamoxifen

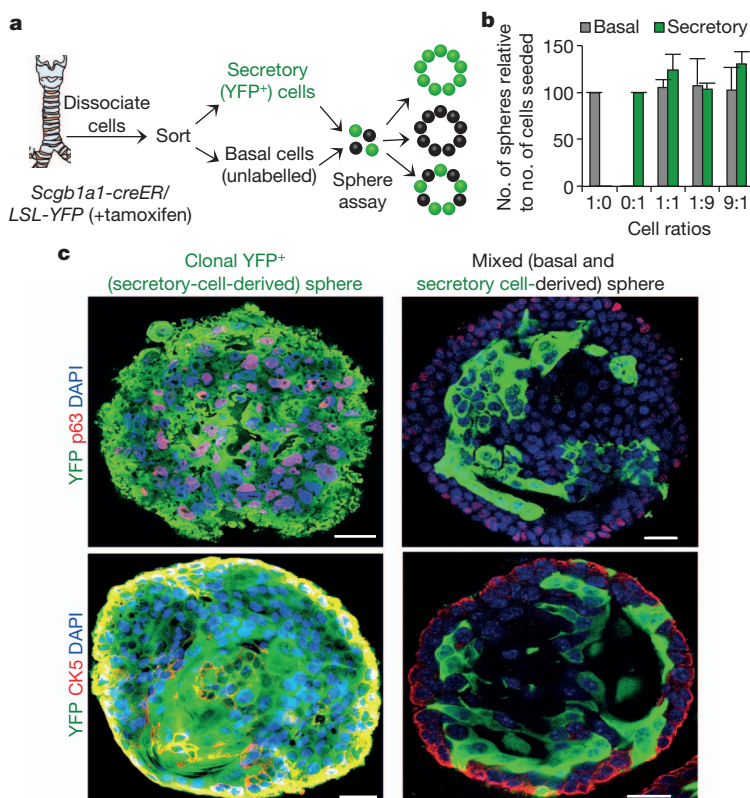
injection (Extended Data Fig. 4). We then performed sphere-forming assays with these pure YFP-labelled secretory cells alone or in combination with pure sorted GSI $\beta$ <sup>4+</sup> unlabelled basal cells in varying proportions. The formation of three possible types of spheres is predicted to occur, including spheres that are comprised of (1) exclusively YFP<sup>+</sup> secretory-cell-derived cells; (2) exclusively unlabelled basal-cell-derived cells; or (3) chimaeric spheres containing YFP<sup>+</sup> secretory-cell-derived and unlabelled basal-cell-derived cells in the same sphere (Fig. 3a). Using our mixed cell populations, we identified a seeding density at which  $99.14\% \pm 0.18$  ( $n = 3$ ) of spheres were either entirely YFP<sup>+</sup> or entirely unlabelled, thus establishing a clonal origin of the spheres. Regardless of the relative ratio of input secretory and basal stem cells cultured together in a well, the aggregate clonal sphere-forming efficiency of each cell type remained constant (Fig. 3b). Immunofluorescence analysis demonstrated that in many spheres originating from a single YFP<sup>+</sup> secretory cell, some YFP<sup>+</sup> cells expressed the basal stem cell markers CK5 and p63, indicating that secretory cell dedifferentiation occurs *ex vivo* (even in the presence of nearby purely stem-cell-derived spheres; Fig. 3c). Thus, we established an *ex vivo* clonal dedifferentiation assay. Of the total spheres derived clonally from YFP<sup>+</sup> secretory cells, 88% showed evidence of dedifferentiation, and in 70.58% of these spheres, more than 10% of cells expressed the stem cell marker CK5. Even when the input proportion of basal cells to secretory cells plated in Matrigel was 100:1, no inhibition of dedifferentiation was observed in the spheres that were clonally derived from secretory cells (data not shown). This suggested the possibility that basal cells in one sphere do not provide a secreted factor to suppress the dedifferentiation of secretory cells in another sphere and that direct contact is required to suppress secretory cell dedifferentiation. Indeed, we observed that in the small fraction ( $0.86\% \pm 0.09$ ;  $n = 3$ ) of chimaerically derived spheres, not a single YFP<sup>+</sup>-labelled secretory cell dedifferentiated and went on to express basal cell markers (Fig. 3c). Therefore a single basal stem cell in direct contact with a secretory cell prevents dedifferentiation. Of note, these findings cannot entirely exclude the possibility that a secreted stem-cell-derived growth factor

can locally suppress secretory cell dedifferentiation. However, they do suggest that an intimate association of basal cells and secretory cells would be required to generate sufficient concentrations of a putative secreted dedifferentiation-suppressing activity.

To provide further confirmation that secretory cells can dedifferentiate into basal cells *ex vivo*, we used mice carrying a doxycycline-inducible basal-cell-specific reporter allele (*CK5-rtTA/tet(O)H2BGFP*) and sorted SSEA-1<sup>+</sup>/H2B-GFP (in which the human histone H2B coding sequence is fused to the GFP coding sequence)<sup>−</sup> secretory cells from their tracheas (Extended Data Fig. 5a, b) and performed a sphere-forming assay. We verified that no GFP<sup>+</sup> cells were present before induction by doxycycline. When doxycycline was administered during the course of sphere formation, the first appearance of H2B-GFP occurred at day 3 of culture, indicating that the secretory cells had been converted into basal-like cells. Immunofluorescence analysis revealed that the resulting H2B-GFP<sup>+</sup> cells also expressed p63 and CK5 (Extended Data Fig. 5c, left panels). The same results were obtained when the sorted SSEA-1<sup>+</sup>/H2B-GFP<sup>−</sup> cells were grown on transwells (Extended Data Fig. 5c, right panels). In addition, secretory cells also dedifferentiated into basal-like cells when we cultured lineage-labelled YFP<sup>+</sup> cells of SCGB1A1-YFP mice on transwell membranes or as spheres (Extended Data Fig. 6a–d). These *ex vivo*-dedifferentiated cells could be serially passaged five times in transwell culture or as spheres and retain their expression of basal stem cell markers and their ability to self-renew (Extended Data Fig. 6d, e). Similarly, cells that had undergone dedifferentiation *in vivo* could also be passaged as stable stem cells (Extended Data Fig. 6f, g). Thus, dedifferentiated secretory cells can stably self-renew.

### Mature secretory cells resist dedifferentiation

To determine whether all secretory cells have the potential to dedifferentiate or whether only a subset of secretory cells is endowed with this capacity, we attempted to subset this class of epithelial cells. To do so, we made use of a transgenic mouse strain that expresses enhanced green fluorescent protein (eGFP) specifically in secretory cells of the



**Figure 3 | Secretory cells dedifferentiate in the absence of basal cells in an *ex vivo* sphere-forming assay.** **a**, Schematic representation of lineage labelling, sorting and *ex vivo* sphere-forming assay. **b**, Quantification of the number of spheres that are either basal-cell-derived (grey bars) and secretory-cell-derived (green bars). x axis, the ratios of basal to secretory cells seeded. y axis, number of spheres formed relative to the number of cells seeded. **c**, Immunostaining for p63 (red; top) or CK5 (red; bottom) in combination with YFP (green).  $n = 3$  (two replicates per condition). Error bars, average  $\pm$  s.e.m. Scale bars, 20  $\mu$ m.



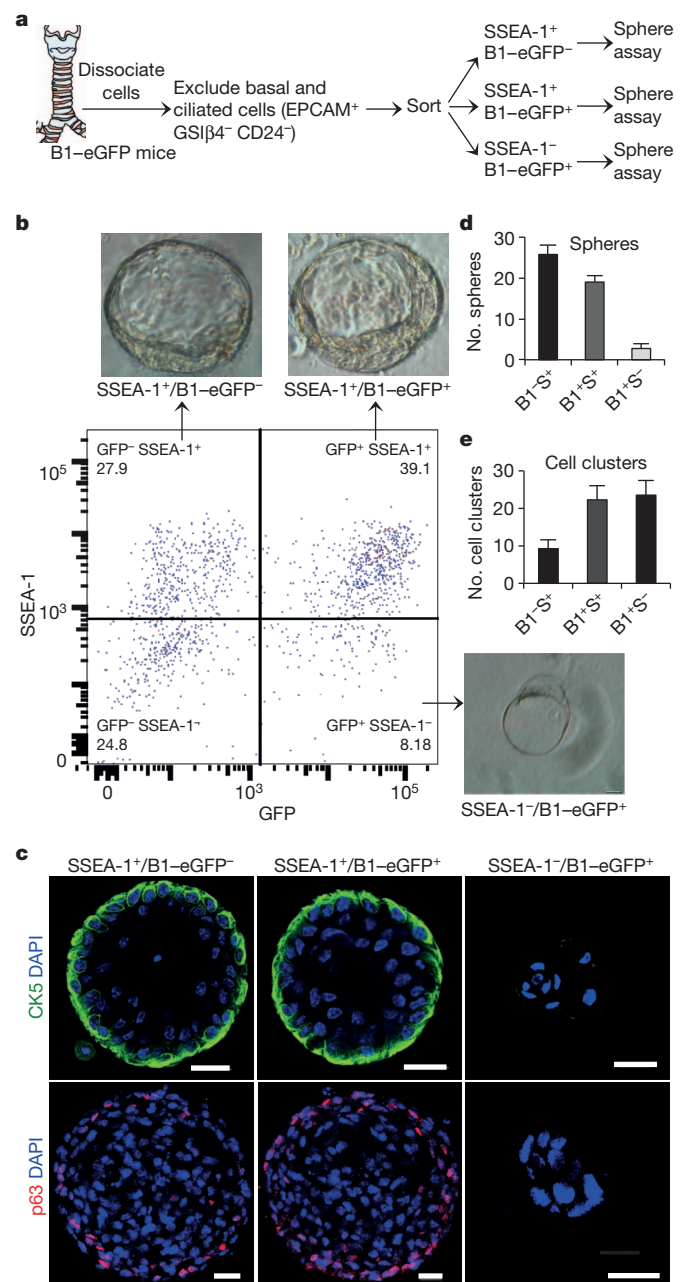
airway epithelium driven by the promoter of the B1 subunit of the vacuolar H(+) -ATPase gene (*Atp6v1b1*); which we reasoned would be associated with mature secretory cells and hereafter refer to as B1-eGFP)<sup>25,26</sup>. Co-immunostaining for SSEA-1 and GFP demonstrated the existence of three subpopulations of secretory cells: SSEA-1<sup>+</sup>/GFP<sup>-</sup>, SSEA-1<sup>+</sup>/GFP<sup>+</sup> and SSEA-1<sup>-</sup>/GFP<sup>+</sup> (Extended Data Fig. 7a). Of note, all GFP<sup>+</sup> cells are SCGB1A1<sup>+</sup> secretory cells and none are CK5<sup>+</sup> basal cells (Extended Data Fig. 7a). To define the cellular hierarchy of these three subsets of cells, we exposed the airway epithelium of B1-eGFP mice to sulphur dioxide (SO<sub>2</sub>) injury. In this injury model, SO<sub>2</sub> causes the complete sloughing of only the suprabasal differentiated cells. The remaining basal stem cells are left intact and start replicating within 24 h to give rise to a mature epithelium within 14 days. We found that single-positive SSEA-1<sup>+</sup> cells appeared first on day 4 and then matured into double-positive SSEA-1<sup>+</sup>/GFP<sup>+</sup> cells on day 6 (Extended Data Fig. 7b) before the formation of any fully mature single-positive B1-eGFP<sup>+</sup> cells evident in the fully mature homeostatic epithelium (Extended Data Fig. 7a, bottom panel, arrowheads). Using B1-eGFP mice, we performed sphere-forming assays with each of the three subsets of secretory cells (Fig. 4a). Intriguingly, all three subsets of cells formed similar large spheres and all these spheres contained basal-like cells (Fig. 4b–d). Interestingly, the sphere-forming ability of the three populations was inversely proportional to the relative maturity of the secretory cell subsets (Fig. 4d). Of note, most of the cell aggregates produced from the most mature SSEA-1<sup>-</sup>/GFP<sup>+</sup> secretory cell subset occurred as small cell clusters instead of spheres (Fig. 4c, e). Furthermore, these cell clusters did not contain CK5- or p63-expressing basal stem cells (Fig. 4c).

### Dedifferentiated cells stably persist

To assess whether dedifferentiated stem-cell-like cells have the ability to self-renew and persist *in vivo*, we generated SCGB1A1-YFP/CK5-DTA mice that possessed lineage-labelled dedifferentiated basal-like cells as above and these mice were then maintained for 2 months before euthanization (Extended Data Fig. 8a). Dedifferentiated YFP<sup>+</sup> CK5<sup>+</sup> cells persisted and continued to represent a sizeable fraction of the stem cell pool ( $9.15\% \pm 0.41$ ;  $n = 3$ ). The relative pool size of dedifferentiated basal-like cells remained stable over the course of 2 months (dedifferentiated basal cells represented 8% of the stem cell pool immediately after dedifferentiation). In addition, triple immunostaining for CK5, YFP and Ki67 revealed that YFP<sup>+</sup> CK5<sup>+</sup> dedifferentiated basal-like cells have the same self-renewal rates as do their normal YFP<sup>-</sup> CK5<sup>+</sup> basal stem cell counterparts (Extended Data Fig. 8b, c).

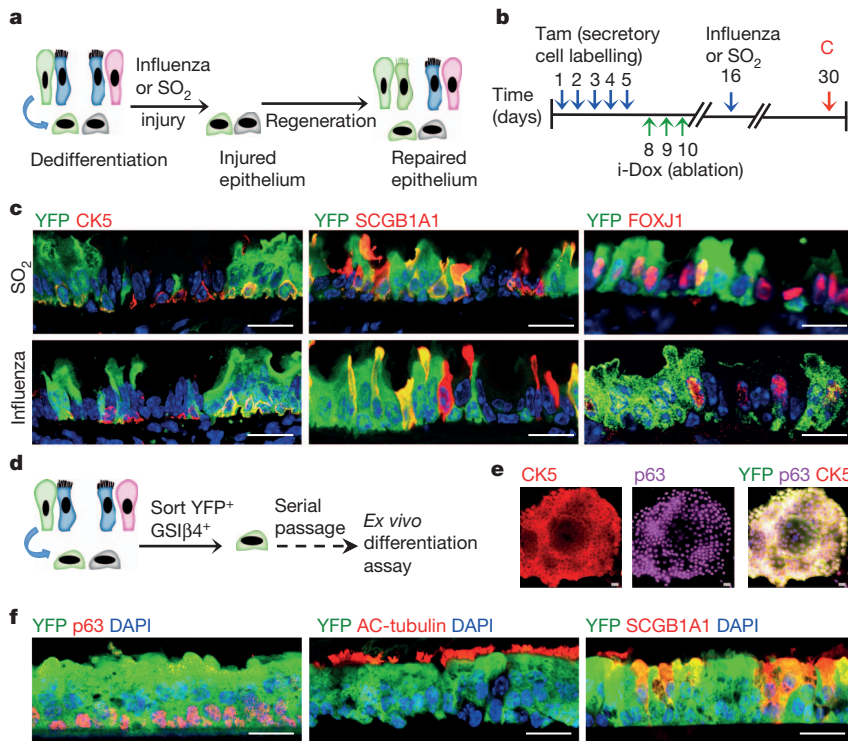
### Dedifferentiated cells are functional stem cells

To assess the functional stem cell capacity of dedifferentiated basal-like cells, we generated SCGB1A1-YFP/CK5-DTA mice that possessed lineage-labelled dedifferentiated basal-like cells and then exposed these animals to two forms of physiologic airway injury (Fig. 5a, b). First, a toxin-induced airway injury with inhaled SO<sub>2</sub> was used to efficiently denude suprabasal cells from the airway epithelium, leaving behind a single layer of basal cells, some of which were derived from labelled secretory cells that had dedifferentiated (marked by YFP) (Extended Data Fig. 9a). The epithelium fully regenerated in 14 days as expected and immunofluorescence analysis for YFP in combination with CK5 (basal cell), SCGB1A1 (secretory cell) and FOXJ1 (ciliated cell) revealed that YFP<sup>+</sup> cells contributed to all three epithelial cell lineages in the form of scattered YFP<sup>+</sup> patches (Fig. 5c, top panels). To further scrutinize the functional potential of our dedifferentiated basal-like cells, we used influenza viral infection as a second physiologic injury model (Extended Data Fig. 9b)<sup>27</sup>. We again observed that dedifferentiated basal-like cells participate in regeneration by giving rise to all three epithelial cell types of the airway (Fig. 5c, bottom panels). Similarly, sorted dedifferentiated cells that were produced either *in vivo* or *ex vivo* could be serially passaged in culture and differentiated into mature airway epithelium (Fig. 5d–f and Extended Data Fig. 10a–c).



**Figure 4 | Dedifferentiation potential of secretory cells is inversely proportional to their maturity.** **a**, Schematic representation of dissociation and sorting of three subsets of secretory cells on the basis of SSEA-1 and GFP expression. **b**, Sorting of secretory cell subsets followed by sphere-forming assay. Representative images of the predominant type of cell aggregates (spheres or cell clusters) from secretory cell subsets. **c**, Immunostaining for CK5 (green) and p63 (red) on cell aggregates. **d, e**, Quantification of the number of spheres (**d**) or cell clusters (**e**) from secretory cell subsets.  $n = 3$  (two replicates per condition). B1, B1-eGFP; S, SSEA-1. Error bars, average  $\pm$  s.e.m. Scale bars, 20  $\mu$ m.

Furthermore, we asked whether individual dedifferentiated basal-like cells are multipotent (that is, able to give rise to ciliated, secretory and basal cells) or unipotent (that is, able to give rise to only one of the cell types). To address this issue, we cultured individual dedifferentiated cells and then performed an air-liquid interface culture using these clonally derived stem cells. Immunofluorescence analyses for basal, secretory and ciliated cell markers revealed that most of the clonally derived basal-like cells (11 out of 13 clones) are multipotent (Extended Data Fig. 10d). Intriguingly, rare clones give rise to only



**Figure 5 | Dedifferentiated cells are functional stem cells both *in vivo* and *ex vivo*.** **a**, Schematic representation of the regeneration of epithelium from dedifferentiated cells after  $\text{SO}_2$ - or influenza-induced injury. **b**, Timeline for the induction of dedifferentiation before infectious or toxic injury followed by tissue collection (C). **c**, Co-labelling of YFP (green) with CK5 (red; left), SCGB1A1 (red; middle) and FOXJ1 (red; right) after  $\text{SO}_2$  (top)- or influenza (bottom)-induced injury. **d**, *Ex vivo* expansion, and differentiation of sorted dedifferentiated cells. **e**, Immunostaining for CK5 (red), p63 (magenta) and YFP (green) on colonies from sorted dedifferentiated cells. **f**, Co-labelling of YFP (green) with p63 or acetylated (AC)-tubulin or SCGB1A1 (red).  $n = 3$  (two replicates per mouse per condition). Scale bars, 20  $\mu\text{m}$ .

ciliated (1 out of 13 clones) or secretory cells (1 out of 13 clones). The potential of single dedifferentiated cells to act as multipotent stem cells suggests that some or most basal stem cells *in vivo* may analogously serve as multipotent stem cells. However, the rare lineage restricted clones we observe may reflect a heterogeneity in the basal cell population *in vivo* that warrants future scrutiny.

## Discussion

Here, using lineage tracing, we have presented evidence that fully differentiated cells can revert into stable and functional basal stem cells. By contrast, in the hair follicle, the committed progeny of skin stem cells do not revert into stem cells when the skin stem cells are ablated<sup>17</sup>. However, in the murine intestinal epithelium, undifferentiated secretory progenitors that are the immediate villin-negative progeny of stem cells can revert back to a stem cell state after injury<sup>13</sup>. Notably, the capacity of fully differentiated villin-positive intestinal secretory cells to revert into stem cells was not assessed. Here we show that fully committed secretory cells respond to stem cell ablation by proliferating and converting into functional epithelial stem cells. Our study points to an alternative cellular mechanism through which tissues can regenerate after stem cell loss. The existence of multiple cellular reservoirs of regenerative capacity may allow a more effective reparative response when one or the other cell type is damaged by a toxic or infectious insult.

The ability of basal stem cells to prevent the dedifferentiation of secretory cells has many implications for tissue biology in general, as stem cells and their progeny can now be seen to reciprocally modulate one another to regulate their relative ratios and thereby overall tissue architecture. In our example, the prevention of secretory cell dedifferentiation occurs through direct contact, even with a single stem cell. This mechanism ensures a precise and local control of epithelial architecture. More generally, it suggests that the reciprocal interactions of stem and committed cells may have been 'designed' to ensure a robust self-organizing property in diverse tissue types.

Furthermore, the ability of differentiated cells to acquire stem cell properties seems to be inversely proportional to the degree of the maturity of the differentiated cells. This is analogous to the results

seen in amphibian nuclear reprogramming in which nuclei from more mature cells were less easily reprogrammed than those of their immature counterparts<sup>28,29</sup>. The capacity to segregate secretory cells according to their maturity and associated ability to resist dedifferentiation provides an ideal *in vivo* experimental model for dissecting the molecular mechanisms through which cells might lock their identity as they mature. Finally, our findings have broad implications for cancer biology, as our results point to an underlying physiologic form of cell plasticity that could be co-opted in the process of tumorigenesis<sup>30–34</sup>. Indeed, some lung cancers seem to be able to resist chemotherapy by using a lineage conversion into a different tumour subtype<sup>35</sup>.

## METHODS SUMMARY

CK5-rtTA<sup>22</sup>, Scgb1a1-creER<sup>19</sup>, tet(O)DTA<sup>23</sup> and B1-eGFP<sup>25,26</sup> mice were described previously. Corn oil or tamoxifen (2 mg per 20 g body weight) were intraperitoneally injected for five consecutive days. Aerosolized PBS or doxycycline was administered by inhalation<sup>24</sup>.  $\text{SO}_2$  injury models have been reported previously<sup>19</sup>. Mice were anesthetized and infected with a sublethal dose of influenza by intranasal inhalation as described previously<sup>27</sup>. Sphere culture and staining was performed as described previously<sup>20</sup>. Immunofluorescence and cell sorting were performed using standard protocols.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 April; accepted 17 October 2013.

Published online 6 November 2013.

- Wolff, G. Entwicklungsphysiologische Studien. I. Die Regeneration der Urodelnlinse. *Arch. Entwickl. Mech. Org.* **1**, 380–390 (1895).
- Brookes, J. P. & Kumar, A. Plasticity and reprogramming of differentiated cells in amphibian regeneration. *Nature Rev. Mol. Cell Biol.* **3**, 566–574 (2002).
- Kai, T. & Spradling, A. Differentiating germ cells can revert into functional stem cells in *Drosophila melanogaster* ovaries. *Nature* **428**, 564–569 (2004).
- Slack, J. M. W. Metaplasia and transdifferentiation: from pure biology to the clinic. *Nature Rev. Mol. Cell Biol.* **8**, 369–378 (2007).
- Kragl, M. et al. Cells keep a memory of their tissue origin during axolotl limb regeneration. *Nature* **460**, 60–65 (2009).
- Lehoczky, J. A., Robert, B. & Tabin, C. J. Mouse digit tip regeneration is mediated by fate-restricted progenitor cells. *Proc. Natl Acad. Sci. USA* **108**, 20609–20614 (2011).



7. Rinkevich, Y., Lindau, P., Ueno, H., Longaker, M. T. & Weissman, I. L. Germ-layer and lineage-restricted stem/progenitors regenerate the mouse digit tip. *Nature* **476**, 409–413 (2011).
8. Sugimoto, K., Gordon, S. P. & Meyerowitz, E. M. Regeneration in plants and animals: dedifferentiation, transdifferentiation, or just differentiation? *Trends Cell Biol.* **21**, 212–218 (2011).
9. Wang, X. *et al.* A luminal epithelial stem cell that is a cell of origin for prostate cancer. *Nature* **461**, 495–500 (2009).
10. Van Keymeulen, A. *et al.* Distinct stem cells contribute to mammary gland development and maintenance. *Nature* **479**, 189–193 (2011).
11. Jopling, C. *et al.* Zebrafish heart regeneration occurs by cardiomyocyte dedifferentiation and proliferation. *Nature* **464**, 606–609 (2010).
12. Jopling, C., Boue, S. & Izpisua Belmonte, J. C. Dedifferentiation, transdifferentiation and reprogramming: three routes to regeneration. *Nature Rev. Mol. Cell Biol.* **12**, 79–89 (2011).
13. van Es, J. H. *et al.* Dll1<sup>+</sup> secretory progenitor cells revert to stem cells upon crypt damage. *Nature Cell Biol.* **14**, 1099–1104 (2012).
14. Dabeva, M. D. *et al.* Liver regeneration and  $\alpha$ -fetoprotein messenger RNA expression in the retrorsine model for hepatocyte transplantation. *Cancer Res.* **58**, 5825–5834 (1998).
15. Brawley, C. & Matunis, E. Regeneration of male germline stem cells by spermatogonial dedifferentiation *in vivo*. *Science* **304**, 1331–1334 (2004).
16. Cheng, J. *et al.* Centrosome misorientation reduces stem cell division during ageing. *Nature* **456**, 599–604 (2008).
17. Hsu, Y., -C., Pasolli, H. A. & Fuchs, E. Dynamics between stem cells, niche, and progeny in the hair follicle. *Cell* **144**, 92–105 (2011).
18. Song, H. *et al.* Functional characterization of pulmonary neuroendocrine cells in lung development, injury, and tumorigenesis. *Proc. Natl Acad. Sci. USA* **109**, 17531–17536 (2012).
19. Rawlins, E. L. *et al.* The role of Scgb1a1<sup>+</sup> Clara cells in the long-term maintenance and repair of lung airway, but not alveolar, epithelium. *Cell Stem Cell* **4**, 525–534 (2009).
20. Rock, J. R. *et al.* Basal cells as stem cells of the mouse trachea and human airway epithelium. *Proc. Natl Acad. Sci. USA* **106**, 12771–12775 (2009).
21. Rock, J. R. & Hogan, B. L. M. Epithelial progenitor cells in lung development, maintenance, repair, and disease. *Annu. Rev. Cell Dev. Biol.* **27**, 493–512 (2011).
22. Diamond, I., Owolabi, T., Marco, M., Lam, C. & Glick, A. Conditional gene expression in the epidermis of transgenic mice using the tetracycline-regulated transactivator/tTA and rTA linked to the keratin 5 promoter. *J. Invest. Dermatol.* **115**, 788–794 (2000).
23. Weber, T. *et al.* Inducible gene expression in GFAP<sup>+</sup> progenitor cells of the SGZ and the dorsal wall of the SVZ—a novel tool to manipulate and trace adult neurogenesis. *Glia* **59**, 615–626 (2011).
24. Tata, P. R. *et al.* Airway specific inducible transgene expression using aerosolized doxycycline. *Am. J. Respir. Cell Mol. Biol.* <http://dx.doi.org/10.1165/rcmb.2012-0412OC> (12 July, 2013).
25. Miller, R. L. *et al.* V-ATPase B1-subunit promoter drives expression of EGFP in intercalated cells of kidney, clear cells of epididymis and airway cells of lung in transgenic mice. *Am. J. Physiol. Cell Physiol.* **288**, C1134–C1144 (2005).
26. Kim, J. K. *et al.* *In vivo* imaging of tracheal epithelial cells in mice during airway regeneration. *Am. J. Respir. Cell Mol. Biol.* **47**, 864–868 (2012).
27. Cho, J. L. *et al.* Enhanced Tim3 activity improves survival after influenza infection. *J. Immunol.* **189**, 2879–2889 (2012).
28. Briggs, R. & King, T. J. Transplantation of living nuclei from blastula cells into enucleated frogs' eggs. *Proc. Natl Acad. Sci. USA* **38**, 455–463 (1952).
29. Gurdon, J. B., Elsdale, T. R. & Fischberg, M. Sexually mature individuals of *Xenopus laevis* from the transplantation of single somatic nuclei. *Nature* **182**, 64–65 (1958).
30. Blanpain, C. Tracing the cellular origin of cancer. *Nature Cell Biol.* **15**, 126–134 (2012).
31. Schwitala, S. *et al.* Intestinal tumorigenesis initiated by dedifferentiation and acquisition of stem-cell-like properties. *Cell* **152**, 25–38 (2013).
32. Friedmann-Morvinski, D. *et al.* Dedifferentiation of neurons and astrocytes by oncogenes can induce gliomas in mice. *Science* **338**, 1080–1084 (2012).
33. Visvader, J. E. Cells of origin in cancer. *Nature* **469**, 314–322 (2011).
34. Goldstein, A. S., Huang, J., Guo, C., Garraway, I. P. & Witte, O. N. Identification of a cell of origin for human prostate cancer. *Science* **329**, 568–571 (2010).
35. Sequist, L. V. *et al.* Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. *Sci. Transl. Med.* **3**, 75ra26 (2011).

**Acknowledgements** The work in this manuscript was supported by a Harvard Stem Cell Institute Seed Grant and a National Institutes of Health-National Heart, Lung, and Blood Institute Early Career Research New Faculty (P30) award (5P30HL101287-02) and a Harvard Stem Cell Institute (HSCI) Junior Investigator Grant to J.R. J.R. is a New York Stem Cell Foundation-Robertson Investigator. We wish to extend our thanks to W. Anderson, Y. Dor, Q. Zhou, A. Brack, J. Galloway and all members of the Rajagopal laboratory for their constructive criticism. We thank the members of the HSCI flow cytometry core facility for help with cell sorting.

**Author Contributions** P.R.T. designed and performed experiments and wrote the manuscript; H.M., A.P.-S., R.Z., M.P., B.M.L. and V.V. performed *ex vivo* experiments; J.L.C. performed influenza infection experiments; A.S. provided *tet(O)DTA* mice and edited the manuscript; S.B. provided *B1-eGFP* mice; B.D.M. reviewed the manuscript; J.R. suggested and co-designed the study and co-wrote the manuscript with P.R.T.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.R. ([jrajagopal@partners.org](mailto:jrajagopal@partners.org)).

## METHODS

**Mouse models.** *CK5-rtTA<sup>22</sup>*, *Scgb1a1-creER<sup>19</sup>*, *tet(O)DTA<sup>23</sup>* and *B1-eGFP<sup>25,26</sup>* mice were described previously. *Rosa26R-eYFP* (*Gt(Rosa)26Sor<sup>tm1(eYFP)Cos</sup>/J*) mice (stock no. 006148) and *Tg(tet(O)HIST1H2BJ/GFP)47Efu/J*; (stock no. 005104) were purchased from The Jackson Laboratory. *CK5-rtTA* females were crossed to *tet(O)DTA* males to generate a double-transgenic mouse (CK5-DTA) that expresses DTA protein upon doxycycline administration. Aerosolized doxycycline or PBS was administered as described previously<sup>24</sup>. For secretory cell lineage tracing after basal cell ablation, we crossed male *Scgb1a1-creER/Rosa26R-YFP* to CK5-DTA female mice to generate quadruple (*SCGB1A1-YFP/CK5-DTA*)-transgenic mice. To label secretory cells, we injected tamoxifen intraperitoneally (2 mg per 20 g body weight) for five consecutive days to induce the Cre-mediated excision of a stop codon and subsequent expression of YFP. Both male and female mice were used for experiments. 6–12-week-old mice were used for experiments. Similar aged mice were used for both control and treated animals. We analysed at least three mice per condition in each experiment. The MGH Subcommittee on Research Animal Care approved animal protocols in accordance with NIH guidelines.

**SO<sub>2</sub> and influenza infection induced injury.** SO<sub>2</sub> injury models have been previously reported<sup>19</sup>. In brief, mice were exposed to 500 p.p.m. of SO<sub>2</sub> for 3 h 40 min. For influenza experiments, influenza A/Puerto Rico/8/34 (PR8) was obtained from Charles River Laboratories International. Mice were anaesthetized and infected with a sublethal dose of influenza by intranasal inhalation as described previously<sup>27</sup>. Mice were killed at day 3 or 14 after infection, and the tracheas were removed and placed in 4% paraformaldehyde (PFA).

**Immunofluorescence, microscopy and cell counting.** Tracheae were dissected and fixed in 4% PFA for 2 h at 4 °C followed by two washes in PBS, and then embedded in OCT. Cryosections (6 µm) were permeabilized with 0.1% Triton X-100 in PBS, blocked in 1% BSA for 30 min at room temperature (27 °C), incubated with primary antibodies for 1 h at room temperature, washed, incubated with appropriate secondary antibodies diluted in blocking buffer for 1 h at room temperature, washed and counterstained with DAPI. Spheres were fixed with 4% PFA on day 9 after plating and washed with PBS and stained as described above.

The following primary antibodies were used: chicken anti-GFP (1:500; GFP-1020, Aves Labs); rabbit anti-Ki67 (1:200; ab15580, Abcam); rat anti-Ki67 (1:200; 14-5698-82, eBioscience); goat anti-SCGB1A1 (1:500; provided by B. Stripp); mouse anti-p63 (1:100; sc-56188, Santa Cruz); mouse anti-tubulin, acetylated (1:100; T6793, Sigma); mouse anti-FOXJ1 (1:500; 14-9965, eBioscience), rabbit anti-NGFR (1:200; ab8875, Abcam); mouse IgM anti-SSEA-1 (1:100; 14-8813-82, eBioscience); hamster anti-T1α (1:50, DSHB) and rabbit anti-cytokeratin 5 (1:1,000; ab53121, Abcam). All secondary antibodies were Alexa Fluor conjugates (488, 594 and 647) and used at 1:500 dilution (Life Technologies).

Images were obtained using Olympus IX81 Inverted microscope (Olympus). Confocal images were obtained with a Nikon A1 confocal laser-scanning microscope with a 40× or 60× oil objective (Nikon CFI Plan APO VC 40× or 60× Oil). Cells were manually counted based on immunofluorescence staining of markers for each of the respective cell types. Cartilage 1 to 9 was used as reference points in all the tracheal samples to count specific cell types on the basis of immunostaining. Serial sections were stained for the antibodies tested and randomly selected slides were used for cell counting. Percentage of basal cells per sphere was calculated

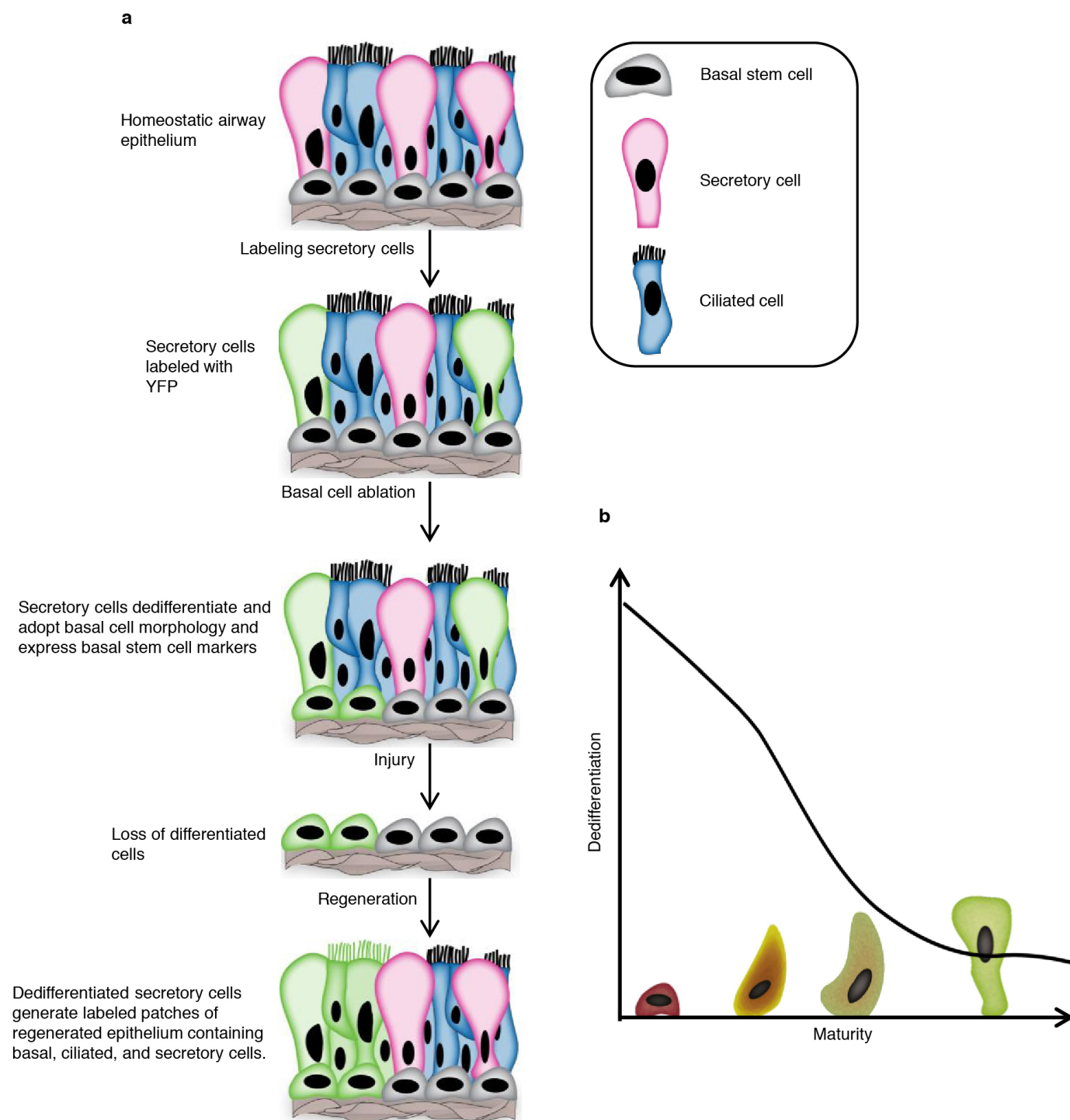
based on CK5 immunostaining. Spheres with less than 2% CK5<sup>+</sup> cells were not included in the quantification.

**Cell dissociation and sorting.** Airway epithelial cells from trachea were dissociated using papain solution. Longitudinal halves of the trachea were cut into five pieces and incubated in papain dissociation solution and incubated at 37 °C for 2 h. After incubation, dissociated tissues were passed through a cell strainer and centrifuged and pelleted at 500g for 5 min. Cell pellets were dispersed and incubated with Ovo-mucoid protease inhibitor (Worthington biochemical Corporation, cat. no. LK003182) to inactivate residual papain activity by incubating on a rocker at 4 °C for 20 min. Cells were then pelleted and stained with EPCAM-PECy7 (1:50; 25-5791-80, eBioscience) or EPCAM-APC (1:50; 17-5791, eBioscience), GSIβ4 (*Griffonia simplicifolia* isolectin beta 4)-Biotin (L2120, Sigma), SSEA-1 eFluor 650NC (1:75, 95-8813-41, eBioscience) and PE anti-mouse CD24 (1:100, 553262, BD Pharmingen) for 30 min in 2.5% FBS in PBS on ice. After washing, cells were sorted on BD FACS Aria (BD Biosciences) using FACS Diva software and analysis was performed using FlowJo (version 10) software.

**Sphere-forming assays and transwell cultures.** Cells were cultured and expanded in complete SAGM (small airway epithelial cell growth medium; Lonza, CC-3118) containing TGF-β/BMP4/WNT antagonist cocktails and 5 µM Rock inhibitor Y-27632 (Selleckbio, S1049). To initiate air-liquid interface cultures, airway basal stem cells were dissociated and seeded onto transwell membranes. After confluence, media was removed from the upper chamber. Mucociliary differentiation was performed with PneumaCult-ALI Medium (StemCell, 05001). Differentiation of airway basal stem cells on an air-liquid interface was followed by directly visualizing beating cilia in real time after 10–14 days. For clonal culture assays, dedifferentiated basal-like cells (GSIβ4<sup>+</sup> YFP<sup>+</sup>) were sorted and plated on collagen-coated plates at low cell density to obtain individual colonies. Individual colonies were isolated using microscopic visualization of single colonies followed by trypsin treatment and pipette-assisted aspiration. Individual colony-derived cells were maintained and expanded separately and used for air-liquid interface culture. Sphere culture was performed as described previously<sup>20</sup>. In brief, 50 µl of 1:4 cold Matrigel/MTEC-plus medium was layered on an 8-well chamber slide (Thermo Scientific, cat. no. 177402) and incubated at 37 °C for 10 min to solidify the Matrigel. Sorted cells were mixed in 2% Matrigel in MTEC-plus and plated on pre-coated 8-well chamber slide at a density of 4,000 cells per well. For mixing assays, sorted cells were seeded at a density of 6,000 cells per well. In each experiment, three independent wells were used for each condition tested. Medium was changed every other day for 9 days. For transwells cultures, cells were suspended in MTEC-plus medium and plated on transwell inserts at a cell density of 6,000 cells per well. Medium was changed every day for 7 days. For basal-cell reporter assays, cells were treated with 1 µg ml<sup>-1</sup> doxycycline either at 24 h after plating or just 24 h before collection. For serial passaging of spheres, medium from culture wells was aspirated, washed with PBS, and then treated with trypsin-EDTA (0.25%) for 2 min. Trypsin was inactivated and dissociated cells were collected and centrifuged at 350g for 3 min at 4 °C. Cells were re-seeded at a 1:20 dilution in Matrigel for the next round of sphere culture.

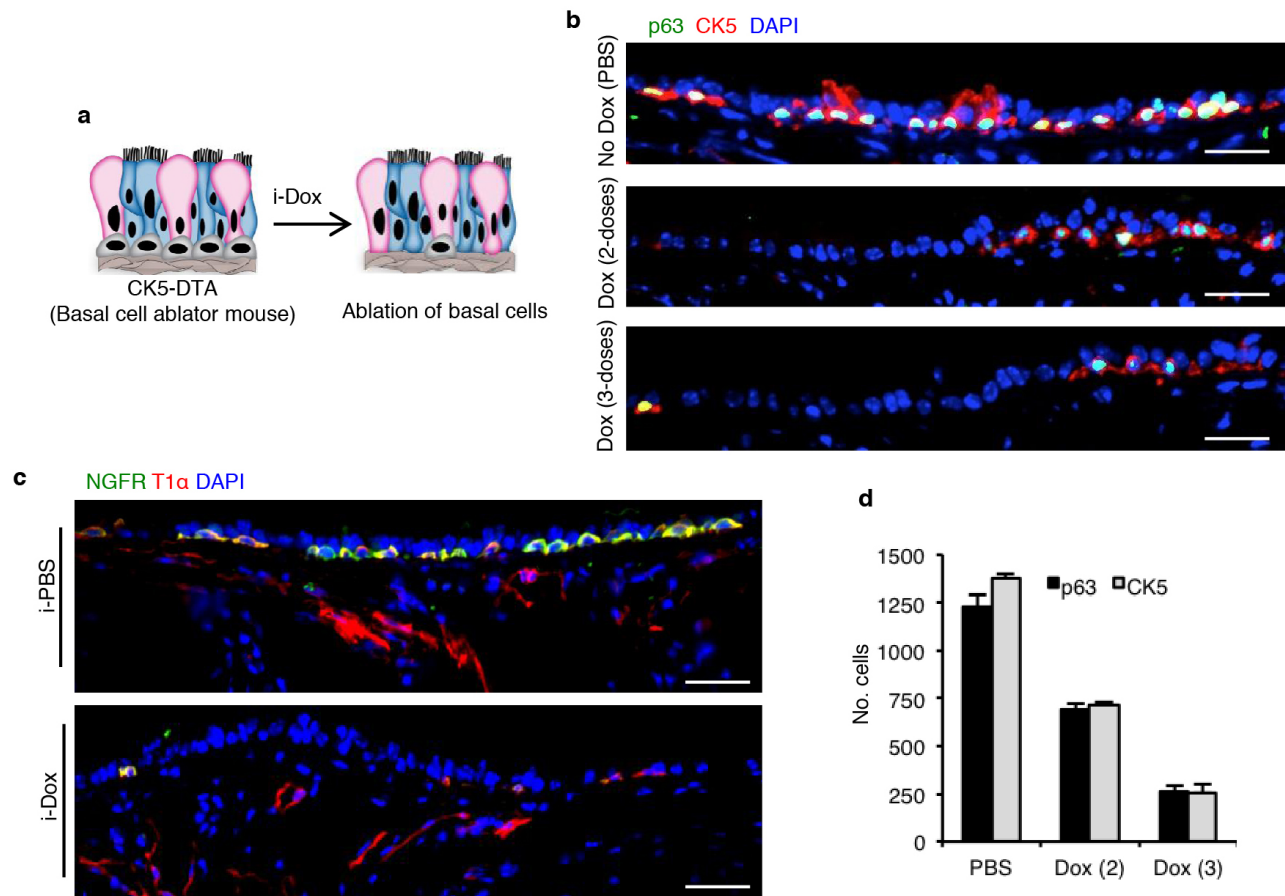
**Statistical analysis.** The standard error of the mean was calculated from the average of at least three independent tracheal samples unless otherwise mentioned. Data were compared among groups using the Student's *t*-test (unpaired, two-tailed). A *P* value of less than 0.05 was considered significant.





**Extended Data Figure 1 | Schematic representation of the dedifferentiation of luminal secretory cells into functional basal stem cells.** **a**, Differentiated luminal secretory cells are labelled with a YFP lineage tag in a homeostatic airway epithelium. Basal stem cells are then ablated using diphtheria toxin. In response, lineage-labelled secretory cells dedifferentiate into cells that morphologically resemble basal cells and express basal stem cell markers. These dedifferentiated basal-like cells respond to physiologically relevant toxic

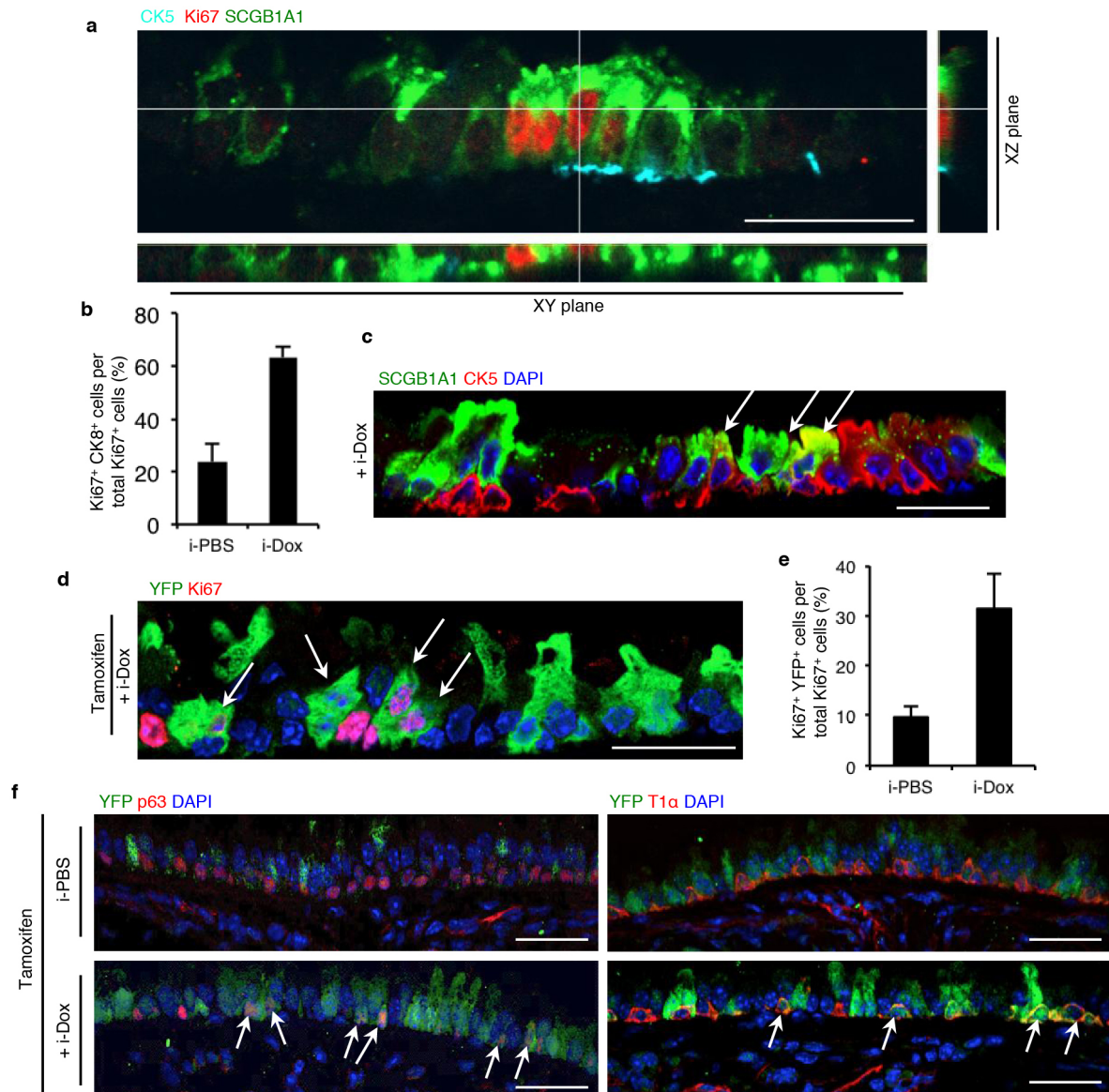
and infectious injury and serve as multipotent stem cells during epithelial regeneration. The inset depicts the different cell types of the airway epithelium. **b**, Graphical representation of the dedifferentiation potential of differentiated basal-like cells. *x* axis represents the maturity of a secretory cell; *y* axis represents the propensity for dedifferentiation to a basal-like cell. The propensity to dedifferentiate is inversely correlated to the maturity of the secretory cell.



**Extended Data Figure 2 | Inhaled doxycycline efficiently ablates basal stem cells of the airway epithelium.** **a**, Schematic representation of basal-cell-specific ablation using i-Dox. **b**, Co-labelling of p63 (green) and CK5 (red) on tracheal sections CK5-DTA mice that received either i-PBS (top) or i-Dox (middle and bottom panels show 2 and 3 doses of i-Dox, respectively). **c**, Co-labelling of NGFR (green) and T1α (red) i-PBS- or i-Dox-treated mice. **d**, Quantification of the number of p63<sup>+</sup> (black bar) and CK5<sup>+</sup> (grey bar) basal

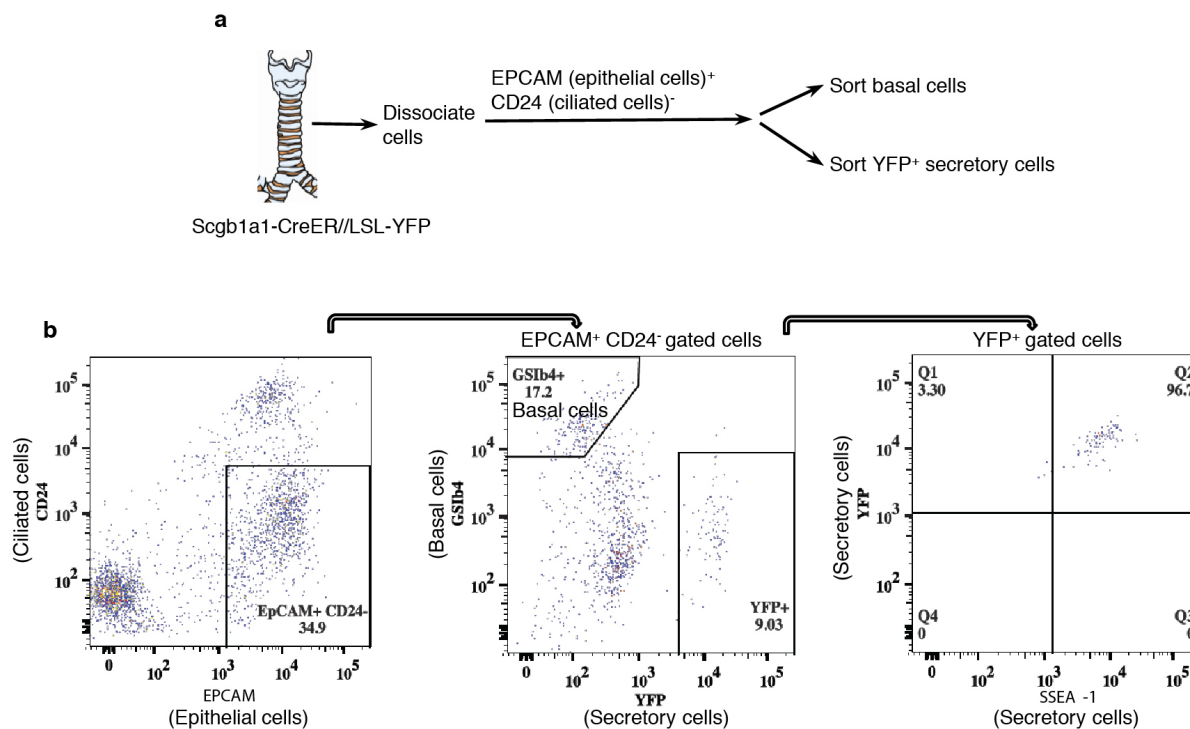
cells from CK5-DTA animals treated with PBS (p63, 1,229 ± 65.45; CK5, 1,376 ± 25.23), two doses of i-Dox (p63, 690 ± 35.13; CK5, 716 ± 12.44) or three doses of i-Dox (p63, 262 ± 29.5; CK5, 255 ± 46.82). *y* axis represents the absolute numbers of basal cells (from three independent tracheal sections). Dox(2) and Dox (3) refer to two and three doses of doxycycline inhalation, respectively. *n* = 3 (two mice per condition). Error bars, average ± s.e.m. Scale bars, 20 μm.





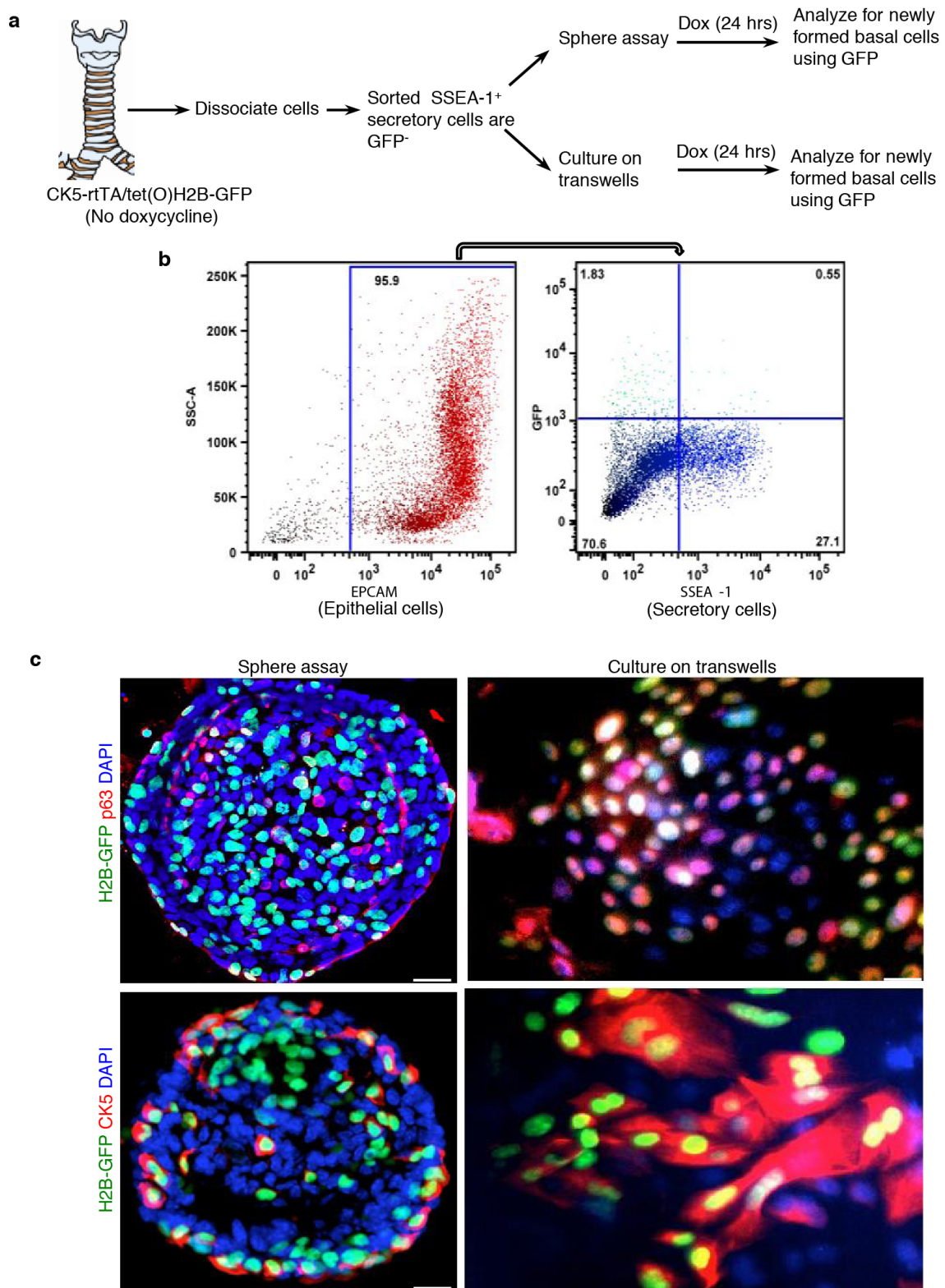
**Extended Data Figure 3 | Secretory cells begin to express proliferation and stem cell markers and undergo dedifferentiation following basal cell ablation.** **a**, Orthogonal confocal optical sections of SCGB1A1 (green), CK5 (cyan) and Ki67 (red) XY and XZ planes are shown to demonstrate the colocalization of Ki67 and SCGB1A1. **b**, Quantification of the percentage of Ki67<sup>+</sup>CK8<sup>+</sup> double-positive cells per total Ki67<sup>+</sup> cells from i-PBS (23.74% ± 6.76)- and i-Dox (63.22% ± 4.14)-treated CK5-DTA mice. **c**, Co-labelling of CK5 and SCGB1A1 in i-Dox-treated CK5-DTA mice. White arrows, double-positive cells. **d**, Immunostaining for YFP (green) and Ki67

(red) on sections from SCGB1A1-YFP/CK5-DTA mice. White arrows indicate YFP<sup>+</sup> Ki67<sup>+</sup> cells. **e**, Quantification of the percentage of YFP<sup>+</sup> Ki67<sup>+</sup> cells per total Ki67<sup>+</sup> cells in SCGB1A1-YFP/CK5-DTA mice that were treated with either i-Dox (31.74% ± 7.15) or i-PBS (9.65% ± 2.12). **f**, Co-labelling of YFP (green) with p63 or T1α (red) on tracheal sections of SCGB1A1-YFP/CK5-DTA mice that were either treated with i-PBS (top) or i-Dox (bottom). White arrows, double-positive cells. *n* = 3 (three mice per condition). Error bars, average ± s.e.m. Scale bars, 20 μm.



**Extended Data Figure 4 | Dissociation and fluorescence activated cell sorting of airway epithelial cells.** **a**, Schematic representation of tracheal epithelial cell dissociation from secretory cell lineage-labelled mice (SCGB1A1-CreER/LSL-YFP) after five doses of tamoxifen. Of the total epithelial cells, EPCAM<sup>+</sup> CD24<sup>-</sup> cells were gated to remove ciliated cells. Then YFP<sup>+</sup>

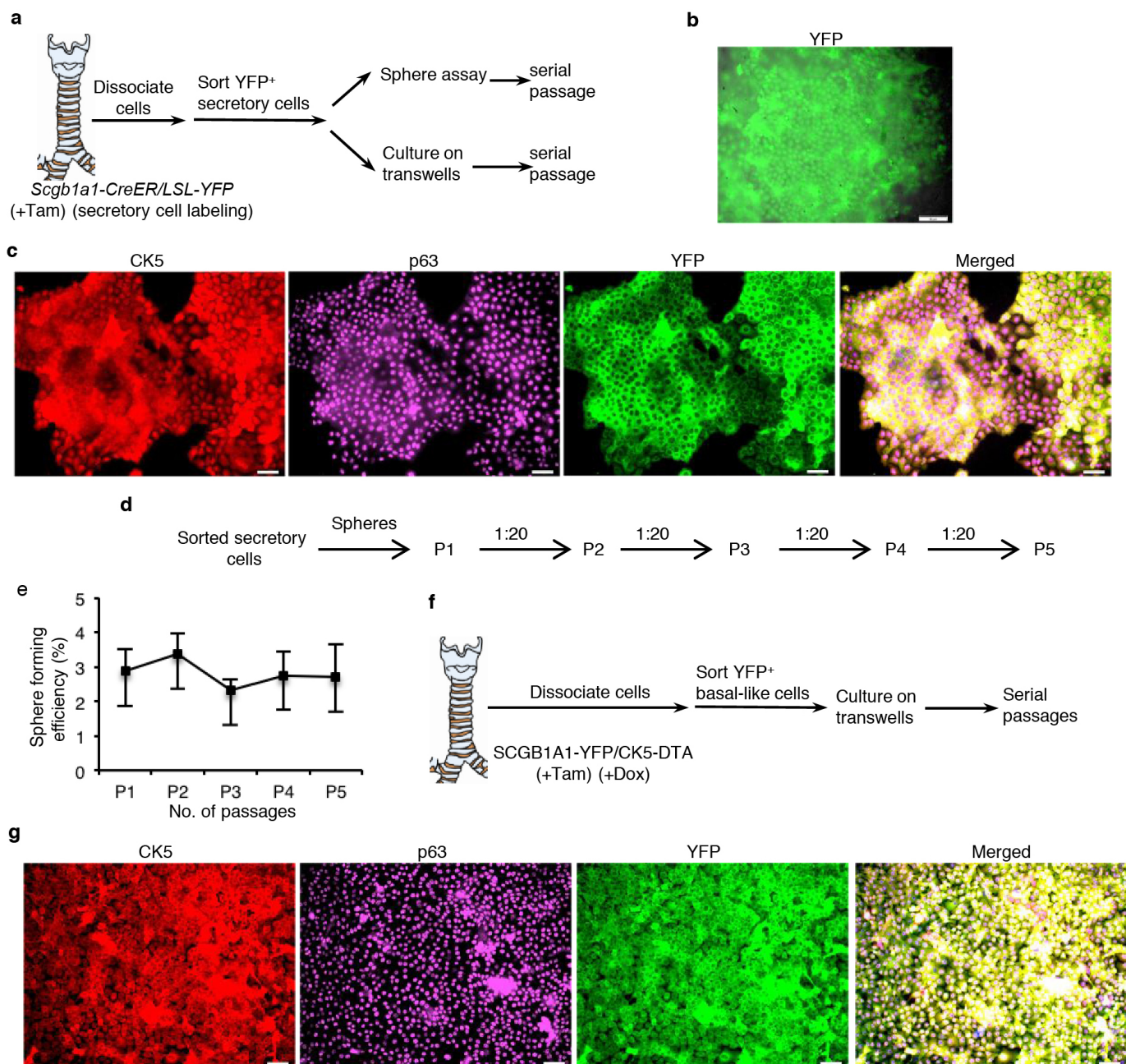
secretory cells and GSIb4<sup>+</sup> basal cells were sorted. **b**, Of the total epithelial cells, EPCAM<sup>+</sup> CD24<sup>-</sup> cells were gated to remove ciliated cells (left). Then, YFP<sup>+</sup> secretory cells were separated from GSIb4<sup>+</sup> basal cells (middle). Sorted YFP<sup>+</sup> cells were also marked by the secretory cell marker SSEA-1 as expected for a pure population of SCGB1A1<sup>+</sup> cells (right).



**Extended Data Figure 5 | Sorted secretory cells dedifferentiate into basal-like self-renewing stem cells upon *ex vivo* culture.** **a**, Schematic representation of tracheal epithelial cell dissociation from basal cell reporter mice (CK5-rtTA/tet(O)H2BGFP) followed by sorting of GFP-SSEA-1<sup>+</sup> secretory cells. Sorted SSEA-1<sup>+</sup> cells were grown as spheres in Matrigel or plated on transwell membranes. Doxycycline was administered and cells were monitored for the initiation of GFP expression. **b**, Fluorescence-activated cell sorting of SSEA-1<sup>+</sup> cells from basal cell reporter mice (CK5-rtTA/tet(O)H2BGFP). Arrows indicate gating windows. EPCAM is a pan-epithelial

marker used to exclude non-epithelial lineages. Sorted SSEA-1<sup>+</sup> secretory cells did not express GFP. **c**, Immunofluorescence staining for p63 (red; top) or CK5 (red; bottom) in combination with H2B-GFP (green; all panels) on sorted secretory cells that were either cultured as Matrigel spheres (left) or on transwells (right). Immunofluorescence analysis confirmed that H2B-GFP<sup>+</sup> cells expressed p63 and CK5, again confirming that secretory cells dedifferentiate in culture. *n* = 3 (two replicates per condition). Scale bars, 20  $\mu$ m.



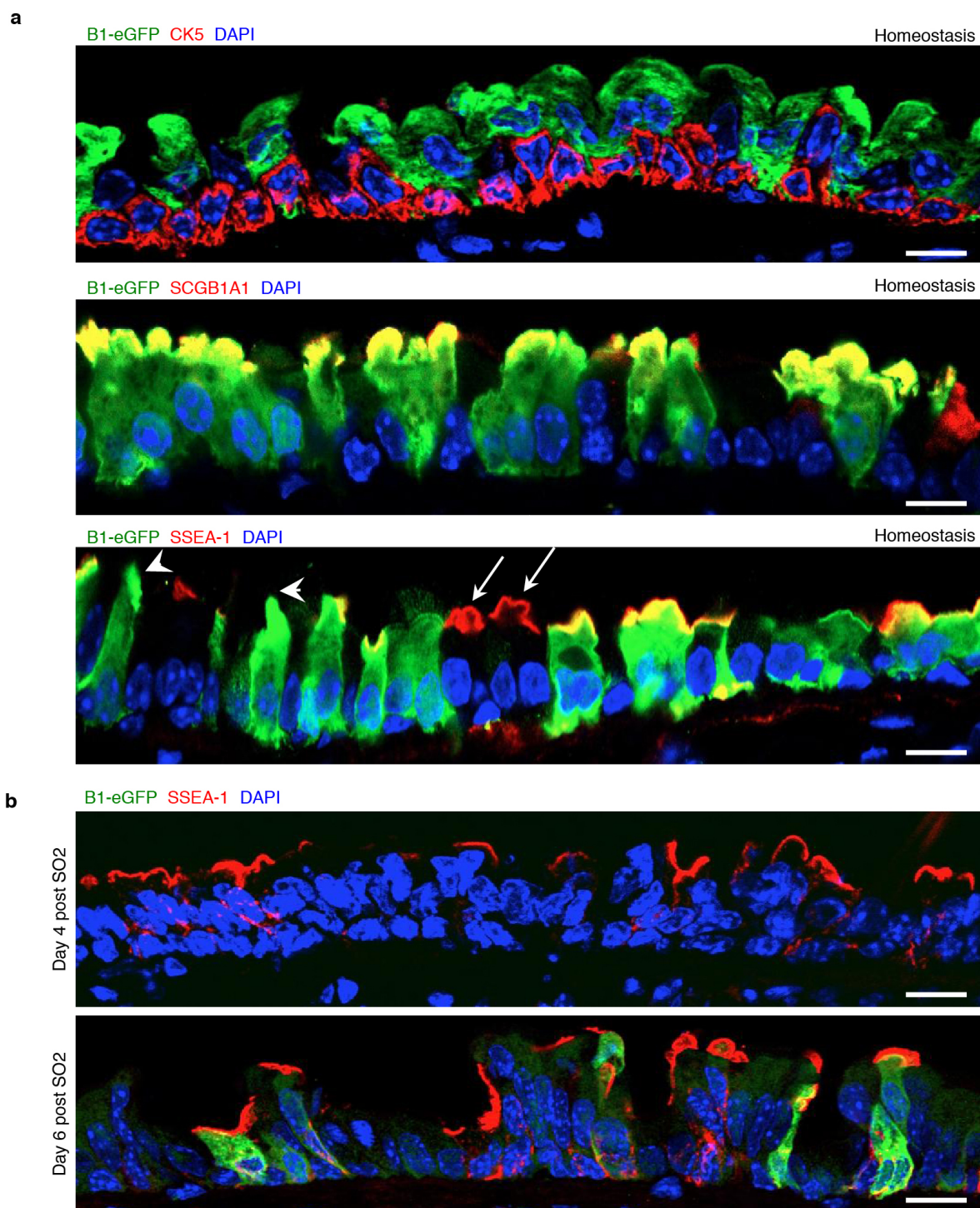


**Extended Data Figure 6 | Sorted lineage-labelled secretory cells undergo dedifferentiation *ex vivo*, express basal stem cell markers, and can be serially passaged, as can secretory cells that underwent dedifferentiation *in vivo*.**

**a**, Schematic representation of secretory cell labelling, sorting and subsequent culturing in Matrigel or on transwell membranes. **b**, Cell colonies obtained from early passage cultures of YFP<sup>+</sup> secretory-cell-derived cells on transwell membranes. **c**, Immunostaining for CK5 (red), p63 (magenta) and YFP (green) on passage-five basal cell colonies from *ex vivo*-dedifferentiated cells. **d**, Schematic showing that YFP<sup>+</sup> secretory-cell-derived spheres from

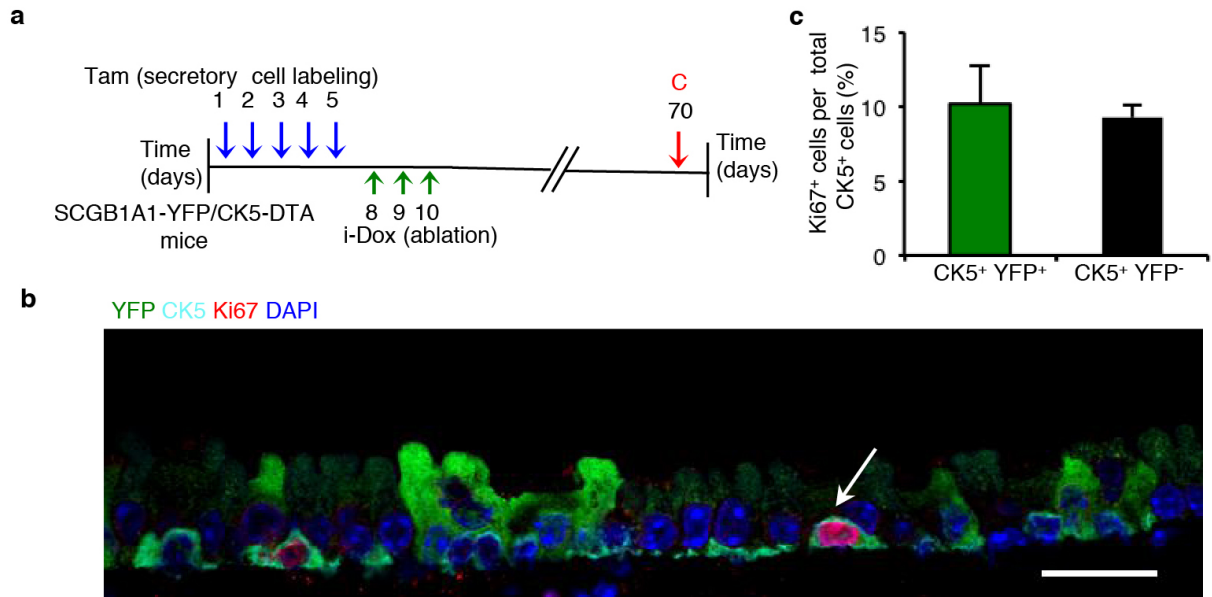
SCGB1A1-CreER/LSL-YFP mice were serially passaged for five generations.

**e**, Quantification of the sphere-forming efficiency: P1 (2.86% ± 0.65), P2 (3.36% ± 0.6), P3 (2.31% ± 0.32), P4 (2.75% ± 0.69) and P5 (2.7% ± 0.94). *x* axis, number of passages. *y* axis, percentage of spheres formed. **f**, Schematic representation of *in vivo* dedifferentiation followed by the sorting and culturing of YFP<sup>+</sup> basal-like cells. **g**, Immunostaining for CK5 (red), p63 (magenta) and YFP (green) on passage-five cell colonies from *in vivo*-dedifferentiated cells. *n* = 3 (two replicates per condition). Error bars, average ± s.e.m. Scale bars, 20 µm.



**Extended Data Figure 7 | B1-eGFP transgenic mice express GFP in mature subsets of secretory cells.** **a**, Co-labelling of GFP (green) with CK5 or SCGB1A1 or SSEA-1 (all in red) on large airways sections derived from adult B1-eGFP transgenic mice at homeostasis. White arrows indicate

SSEA-1<sup>+</sup> B1-eGFP<sup>-</sup>, whereas white arrowheads point to cells that are B1-eGFP<sup>+</sup> SSEA-1<sup>-</sup> (bottom). **b**, B1-eGFP trachea were stained for GFP (green) and SSEA-1 (red) on day 4 and day 6 after SO<sub>2</sub>-induced injury.  $n = 3$  (two mice per condition per time point). Scale bars, 20  $\mu$ m.



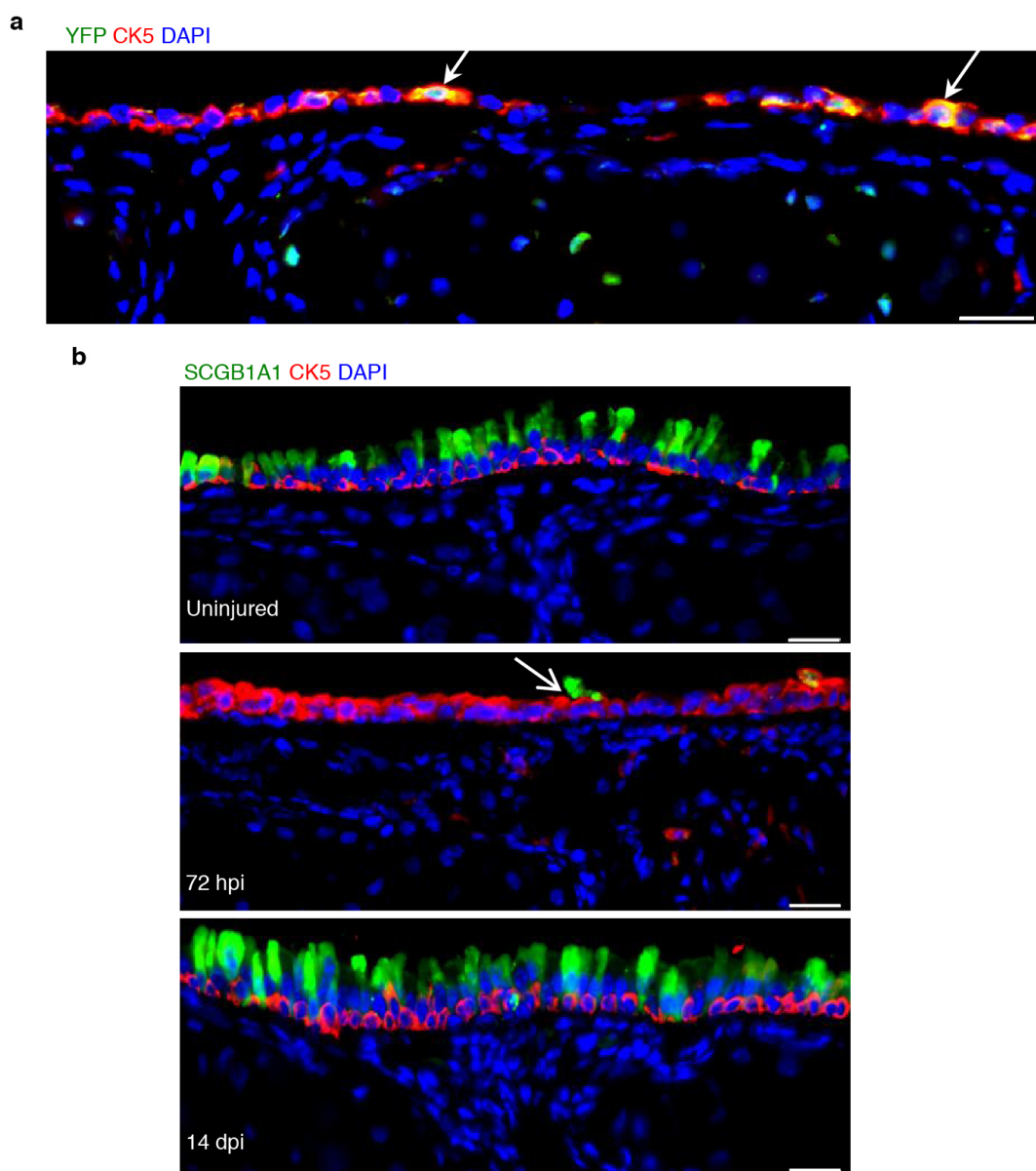
**Extended Data Figure 8 | Dedifferentiated basal-like stem cells are stable and self-renew to the same degree as endogenous basal stem cells.**

**a**, Schematic representation of the dedifferentiation protocol to assess the ability of basal-like stem cells to persist and self-renew for 2 months.

**b**, Immunofluorescence staining for YFP (green) in combination with CK5 (cyan) or Ki67 (red) on sections from SCGB1A1-YFP/CK5-DTA mice 2 months after basal cell ablation. **c**, Quantification of the percentage of

proliferating dedifferentiated CK5<sup>+</sup> YFP<sup>+</sup> (10.16% ± 2.57) (green bar) and wild-type CK5<sup>+</sup> YFP<sup>-</sup> (9.25% ± 0.79) (black bar) stem cells out of the total CK5<sup>+</sup> stem cell population in the large airways of SCGB1A1-YFP/CK5-DTA mice 2 months after basal cell ablation. The white arrow points to a proliferating dedifferentiated basal-like stem cell. *n* = 3 (two mice per condition). Error bars, average ± s.e.m. Scale bar, 20 μm.

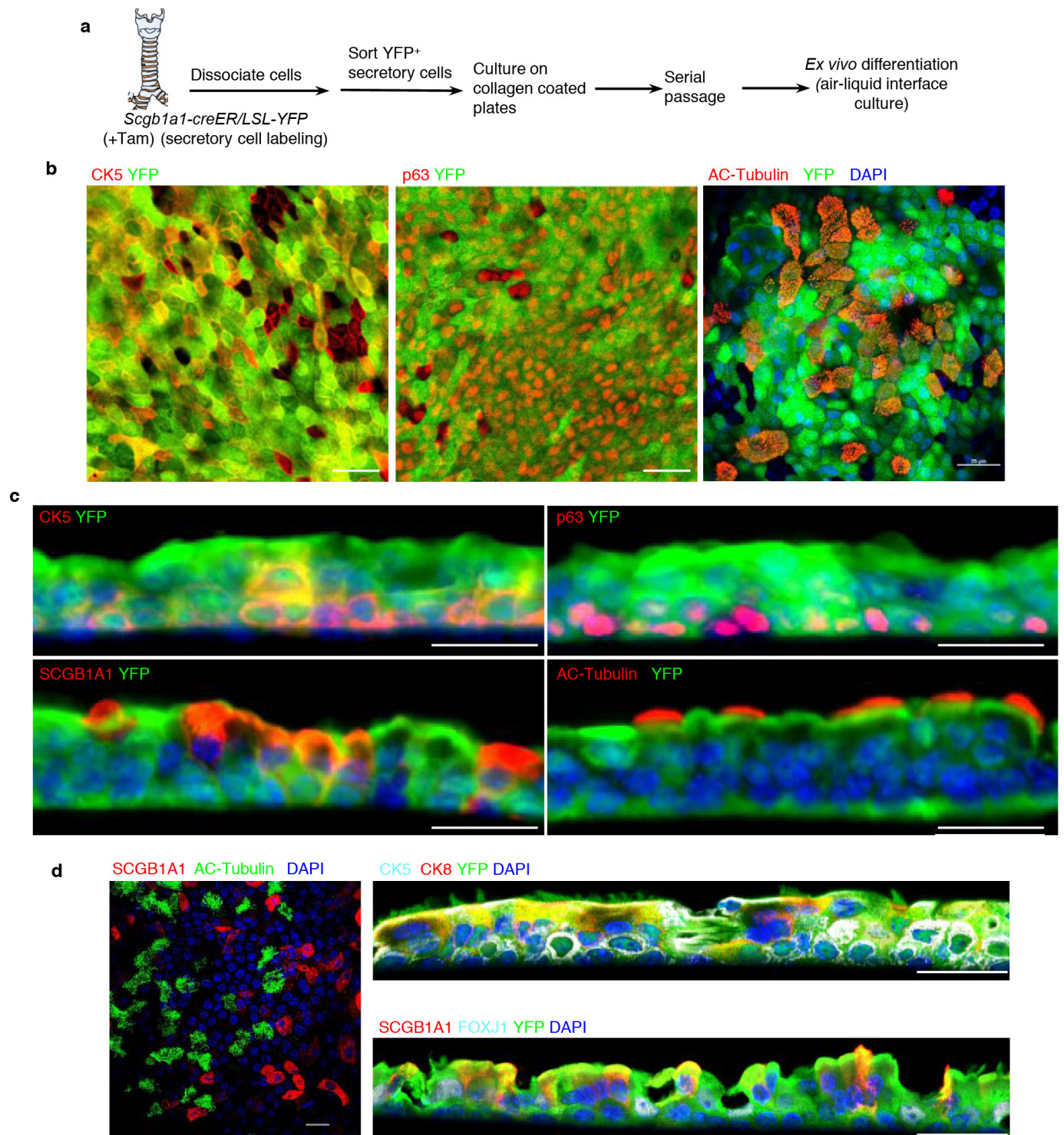




**Extended Data Figure 9 | SO<sub>2</sub>-and influenza-induced injury efficiently removes the suprabasal cells of the airway epithelium.**

**a**, Immunofluorescence staining for CK5 (red) and YFP (green) in SCGB1A1–YFP/CK5-DTA airway epithelium 24 h after SO<sub>2</sub> inhalation. Only a single layer of basal cells persists after injury. White arrows, CK5<sup>+</sup> YFP<sup>+</sup> double-positive

cells (yellow). **b**, Immunofluorescence analysis of basal cells (CK5, red) and secretory cells (SCGB1A1, green) from uninjured (top), 72 hours post-infection (h.p.i.; middle) and 14 days post-infection (d.p.i.; bottom). White arrow, secretory cell debris (in green).  $n = 3$  (two replicates per condition). Scale bars, 20  $\mu\text{m}$ .



**Extended Data Figure 10 | Ex vivo-dedifferentiated cells can be clonally expanded and give rise to a complete airway epithelium in air-liquid interface cultures.** **a**, Schematic representation of the ex vivo dedifferentiation of secretory cells, serial passage and differentiation. **b**, Whole-mount immunostaining for YFP (green) in combination with CK5 or p63 or acetylated (AC)-tubulin (red). **c**, Co-labelling of YFP (green) with CK5 or p63 or

SCGB1A1 or acetylated-tubulin (red) on differentiated epithelium derived from serially passaged dedifferentiated basal-like cells. **d**, Whole-mount immunofluorescence staining for SCGB1A1 (red) and acetylated-tubulin (green) on clonally derived epithelium (left). Co-labelling of YFP (green) with CK5 (white) and CK8 (red; top) or SCGB1A1 and FOXJ1 (bottom) on cross sections of clonally derived epithelium. Scale bars, 20  $\mu$ m.

# Therapeutic efficacy of potent neutralizing HIV-1-specific monoclonal antibodies in SHIV-infected rhesus monkeys

Dan H. Barouch<sup>1,2</sup>, James B. Whitney<sup>1</sup>, Brian Moldt<sup>3</sup>, Florian Klein<sup>4</sup>, Thiago Y. Oliveira<sup>4</sup>, Jinyan Liu<sup>1</sup>, Kathryn E. Stephenson<sup>1</sup>, Hui-Wen Chang<sup>1</sup>, Karthik Shekhar<sup>5</sup>, Sanjana Gupta<sup>5</sup>, Joseph P. Nkolola<sup>1</sup>, Michael S. Seaman<sup>1</sup>, Kaitlin M. Smith<sup>1</sup>, Erica N. Borducchi<sup>1</sup>, Crystal Cabral<sup>1</sup>, Jeffrey Y. Smith<sup>1</sup>, Stephen Blackmore<sup>1</sup>, Srisowmya Sanisetty<sup>1</sup>, James R. Perry<sup>1</sup>, Matthew Beck<sup>6</sup>, Mark G. Lewis<sup>7</sup>, William Rinaldi<sup>8</sup>, Arup K. Chakraborty<sup>2,5</sup>, Pascal Poignard<sup>3</sup>, Michel C. Nussenzweig<sup>4,9\*</sup> & Dennis R. Burton<sup>2,3\*</sup>

**Human immunodeficiency virus type 1 (HIV-1)-specific monoclonal antibodies with extraordinary potency and breadth have recently been described. In humanized mice, combinations of monoclonal antibodies have been shown to suppress viraemia, but the therapeutic potential of these monoclonal antibodies has not yet been evaluated in primates with an intact immune system. Here we show that administration of a cocktail of HIV-1-specific monoclonal antibodies, as well as the single glycan-dependent monoclonal antibody PGT121, resulted in a rapid and precipitous decline of plasma viraemia to undetectable levels in rhesus monkeys chronically infected with the pathogenic simian-human immunodeficiency virus SHIV-SF162P3. A single monoclonal antibody infusion afforded up to a 3.1 log decline of plasma viral RNA in 7 days and also reduced proviral DNA in peripheral blood, gastrointestinal mucosa and lymph nodes without the development of viral resistance. Moreover, after monoclonal antibody administration, host Gag-specific T-lymphocyte responses showed improved functionality. Virus rebounded in most animals after a median of 56 days when serum monoclonal antibody titres had declined to undetectable levels, although, notably, a subset of animals maintained long-term virological control in the absence of further monoclonal antibody infusions. These data demonstrate a profound therapeutic effect of potent neutralizing HIV-1-specific monoclonal antibodies in SHIV-infected rhesus monkeys as well as an impact on host immune responses. Our findings strongly encourage the investigation of monoclonal antibody therapy for HIV-1 in humans.**

A series of broad and potent HIV-1 Env-specific monoclonal antibodies have recently been isolated<sup>1,2</sup> and have been shown to target the CD4 binding site<sup>3–7</sup>, the V1/V2 loops<sup>8,9</sup>, the V3/V4 loops and N332 glycans<sup>10–13</sup>, and the membrane proximal external region<sup>14</sup>. Previous studies in humanized mice and humans using the earlier generation of HIV-1 Env-specific monoclonal antibodies suggested that the therapeutic potential of monoclonal antibodies would be severely limited by the rapid emergence of viral escape mutations in the context of diverse virus swarms<sup>15–17</sup>. However, cocktails of three or five of the new generation of more potent monoclonal antibodies targeting multiple epitopes have recently been shown to suppress HIV-1 replication in humanized mice<sup>18,19</sup>.

## Therapeutic efficacy of monoclonal antibody cocktails

To evaluate the therapeutic potential of broad and potent HIV-1-specific monoclonal antibodies in primates with an intact immune system, we infused cocktails of monoclonal antibodies as well as single monoclonal antibodies into chronically SHIV-infected rhesus monkeys. We focused on the N332 glycan-dependent monoclonal antibody PGT121 (ref. 10) and the CD4-binding-site-specific monoclonal antibodies 3BNC117 (ref. 6) and b12 (ref. 20). In the first study, we used eight Indian origin adult rhesus monkeys (*Macaca mulatta*) that did not express the class I alleles *Mamu-A\*01*, *Mamu-B\*08* and *Mamu-B\*17* and that were

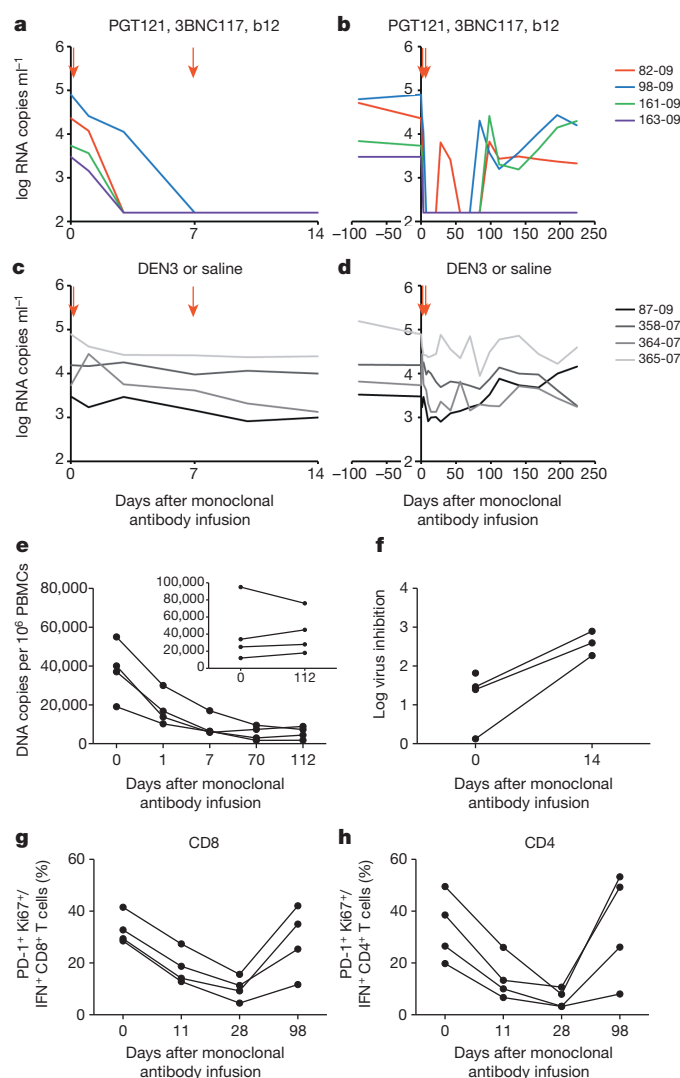
infected intrarectally with the pathogenic virus SHIV-SF162P3 9 months before the monoclonal antibody infusions. These animals exhibited chronic set-point viral loads of 3.4–4.9 log RNA copies ml<sup>-1</sup> with clinical disease progression and reduced CD4<sup>+</sup> T-lymphocyte counts. We performed two intravenous monoclonal antibody infusions on day 0 and day 7 with 10 mg kg<sup>-1</sup> of each of PGT121, 3BNC117 and b12 ( $n = 4$ ); or with 30 mg kg<sup>-1</sup> of the isotype-matched control monoclonal antibody DEN3 ( $n = 1$ ) or saline ( $n = 3$ ).

After the initial monoclonal antibody infusion, we observed rapid and precipitous declines of plasma viral loads to undetectable levels by day 7 in 4 of 4 monkeys (Fig. 1a). Virological control persisted for 84 to 98 days in animals 82-09, 98-09 and 161-09 (Fig. 1b). After viral rebound, sequence analysis<sup>18,21</sup> showed no N332 or other characteristic escape mutations (Supplementary Information), and viral rebound correlated with the decline of serum monoclonal antibody titres to undetectable levels <1 µg ml<sup>-1</sup> (Extended Data Fig. 1). Monkey 82-09 exhibited transient viraemia on day 28 (Fig. 1b), which correlated with the decline of serum monoclonal antibody titres to undetectable levels (Extended Data Fig. 1), but this animal then spontaneously re-controlled viral replication until day 98. Monkey 163-09, which had the lowest baseline viral load of 3.4 log RNA copies ml<sup>-1</sup> before the monoclonal antibody infusion, exhibited long-term virological control for over

<sup>1</sup>Center for Virology and Vaccine Research, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215, USA. <sup>2</sup>Ragon Institute of MGH, MIT, and Harvard, Cambridge, Massachusetts 02139, USA. <sup>3</sup>The Scripps Research Institute, La Jolla, California 92037, USA. <sup>4</sup>The Rockefeller University, New York, New York 10065, USA. <sup>5</sup>Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>6</sup>New England Primate Research Center, Southborough, Massachusetts 01776, USA. <sup>7</sup>Bioqual, Inc., Rockville, Maryland 20852, USA. <sup>8</sup>Alpha Genesis, Inc., Yemassee, South Carolina 29945, USA. <sup>9</sup>Howard Hughes Medical Institute, New York, New York 10065, USA.

\*These authors contributed equally to this work.





**Figure 1 | Therapeutic efficacy of the triple PGT121, 3BNC117 and b12 monoclonal antibody cocktail.** **a, b**, Plasma viral RNA (log copies  $\text{ml}^{-1}$ ) in rhesus monkeys chronically infected with SHIV-SF162P3 after infusions of PGT121, 3BNC117 and b12 on day 0 and day 7 (arrows) for 14 days (**a**) and 224 days (**b**). **c, d**, Plasma viral RNA in rhesus monkeys chronically infected with SHIV-SF162P3 after infusions with the control monoclonal antibody DEN3 (87-09) or saline on day 0 and day 7 (arrows) for 14 days (**c**) and 224 days (**d**). **e**, Proviral DNA (copies per  $10^6$  peripheral blood mononuclear cells (PBMCs)) in the monkeys that received the therapeutic monoclonal antibody cocktail or controls (inset). **f**, Log inhibition of viral replication in  $\text{CD8}^+$  T-lymphocyte-dependent virus suppression assays after monoclonal antibody infusion. One animal had no recoverable virus at week 2. **g, h**,  $\text{PD-1}^+ \text{Ki67}^+$  expression on  $\text{Gag-specific CD8}^+$  (**g**) and  $\text{CD4}^+$  (**h**) T lymphocytes after monoclonal antibody infusion.

200 days despite the absence of detectable serum monoclonal antibody titres after day 70 (Fig. 1b). Proviral DNA in peripheral blood also declined rapidly by tenfold in the monkeys that received the monoclonal antibodies (Fig. 1e). Virological control was not observed in the monkeys that received DEN3 or saline (Fig. 1c, d), and viral loads on day 14 were significantly lower in the monoclonal-antibody-treated monkeys than in the controls ( $P = 0.02$ , Mann–Whitney  $U$ -test).

As expected, serum neutralizing antibody titres<sup>22</sup> to the SHIV-SF162P3 challenge virus increased markedly after the monoclonal antibody administration and then declined over time (Extended Data Fig. 2). After clearance of the monoclonal antibodies, neutralizing antibody titres to SHIV-SF162P3 as well as to the related neutralization-sensitive virus SHIV-SF162P4 remained slightly higher than baseline titres (Extended Data Fig. 2). The magnitude of total  $\text{Gag-specific CD8}^+$

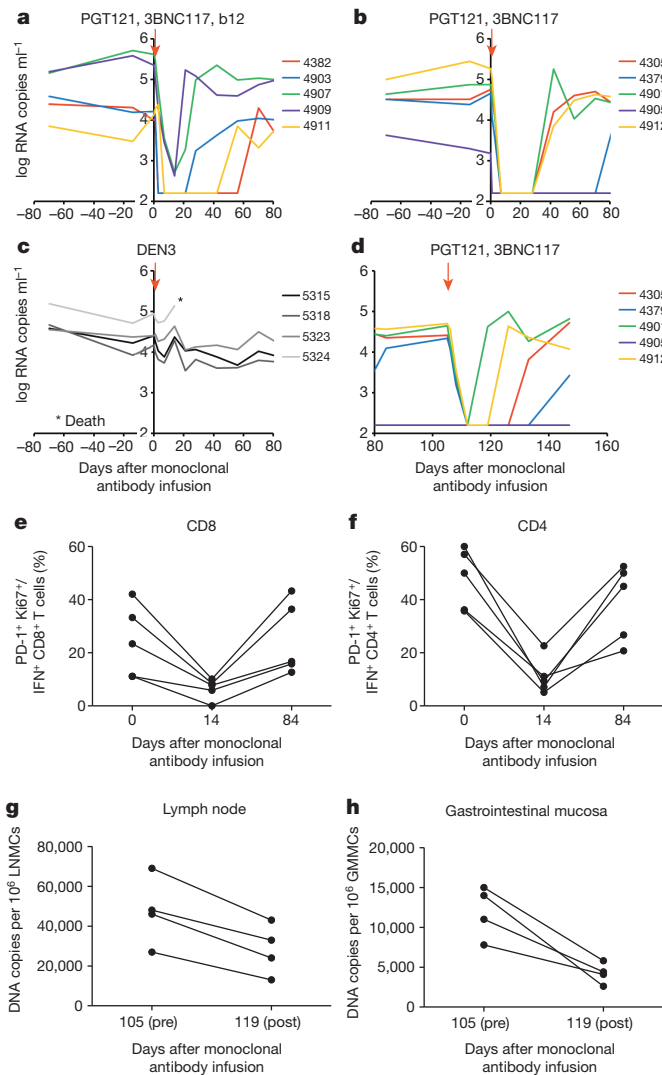
and  $\text{CD4}^+$  T-lymphocyte responses<sup>23,24</sup> was not detectably modulated after monoclonal antibody administration (Extended Data Fig. 3). However, by day 28, we observed threefold and fivefold reductions, respectively, in the percentage of  $\text{Gag-specific CD8}^+$  and  $\text{CD4}^+$  T lymphocytes that expressed the exhaustion and activation markers PD-1 and Ki67 (Fig. 1g, h;  $P = 0.02$ ). Moreover,  $\text{CD8}^+$  T lymphocytes from these animals exhibited increased functional capacity to suppress virus replication after monoclonal antibody infusion<sup>25</sup> (Fig. 1f;  $P = 0.03$ ). These data indicate that monoclonal antibody administration not only exerted direct antiviral effects but also improved host immune responses.

We next investigated the therapeutic efficacy of a single infusion of the cocktail of three monoclonal antibodies as well as a combination of only two monoclonal antibodies. Fourteen rhesus monkeys infected with SHIV-SF162P3 9 months before the monoclonal antibody infusion with chronic set-point viral loads of  $3.2$ – $5.6$  log RNA copies  $\text{ml}^{-1}$  received a single infusion on day 0 with  $10 \text{ mg kg}^{-1}$  of each of the monoclonal antibodies PGT121, 3BNC117 and b12 ( $n = 5$ ); PGT121 and 3BNC117 ( $n = 5$ ); or the isotype-matched control monoclonal antibody DEN3 ( $n = 4$ ). We observed rapid virological control to undetectable levels by day 7 in 3 of 5 animals that received the cocktail of three monoclonal antibodies and in 5 of 5 animals that received only PGT121 and 3BNC117 (Fig. 2a, b). The two animals that failed to achieve complete virological suppression had the highest baseline plasma viral loads of  $5.4$  and  $5.6$  log RNA copies  $\text{ml}^{-1}$  before the monoclonal antibody infusion and exhibited  $2.8$  and  $2.9$  log declines, respectively, in plasma viraemia before rapid viral rebound on day 21 (monkeys 4907, 4909; Fig. 2a). No characteristic viral escape mutations were detected after viral rebound (Supplementary Information). The animals that suppressed viral loads to undetectable levels exhibited up to a  $3.1$  log decline of plasma viral RNA copies  $\text{ml}^{-1}$  by day 7 (monkey 4912; Fig. 2b). Viral rebound occurred in most animals between day 28 and day 84 (Fig. 2a, b) and was associated with declines of serum monoclonal antibody titres to undetectable levels (Extended Data Fig. 4). The animal with the lowest baseline viral load of  $3.2$  log RNA copies  $\text{ml}^{-1}$  exhibited long-term virological control for over 100 days (monkey 4905; Fig. 2b), and viral loads on day 14 were significantly lower in both groups of monoclonal-antibody-treated monkeys than in the controls ( $P = 0.01$ ). One control animal (monkey 5324; Fig. 2c) was euthanized for progressive clinical AIDS and opportunistic infections during the course of this experiment. PGT121 and 3BNC117 administration also resulted in threefold and fourfold reductions in the percentage of  $\text{PD-1}^+ \text{Ki67}^+$   $\text{Gag-specific CD8}^+$  and  $\text{CD4}^+$  T lymphocytes, respectively (Fig. 2e, f).

To confirm that viral rebound was not associated with the development of viral resistance to the monoclonal antibodies, we performed a second infusion of monoclonal antibodies on day 105 in the monkeys that received PGT121 and 3BNC117. Viral re-suppression was observed in 4 of 4 animals after the second monoclonal antibody infusion (Fig. 2d and Supplementary Information). However, virological control appeared less durable and serum monoclonal antibody titres were lower after the second monoclonal antibody infusion as compared with the first monoclonal antibody infusion (Fig. 2b, d and Extended Data Fig. 4), presumably as a result of monkey anti-human antibody responses that developed after the first monoclonal antibody administration (Extended Data Fig. 5). Nevertheless, we assessed the impact of the second monoclonal antibody infusion on proviral DNA in various tissue compartments<sup>26,27</sup> and observed a twofold decline in lymph nodes ( $P = 0.1$ ; Fig. 2g) and a threefold decline in gastrointestinal mucosa ( $P = 0.02$ ; Fig. 2h) 14 days after monoclonal antibody re-administration. These data indicate that the potent monoclonal antibodies not only suppressed viraemia but also reduced proviral DNA in tissues without the generation of viral resistance.

### Therapeutic efficacy of single monoclonal antibodies

Although the cloned SHIV-SF162P3 pseudovirus is highly sensitive to 3BNC117, we observed that our particular SHIV-SF162P3 challenge



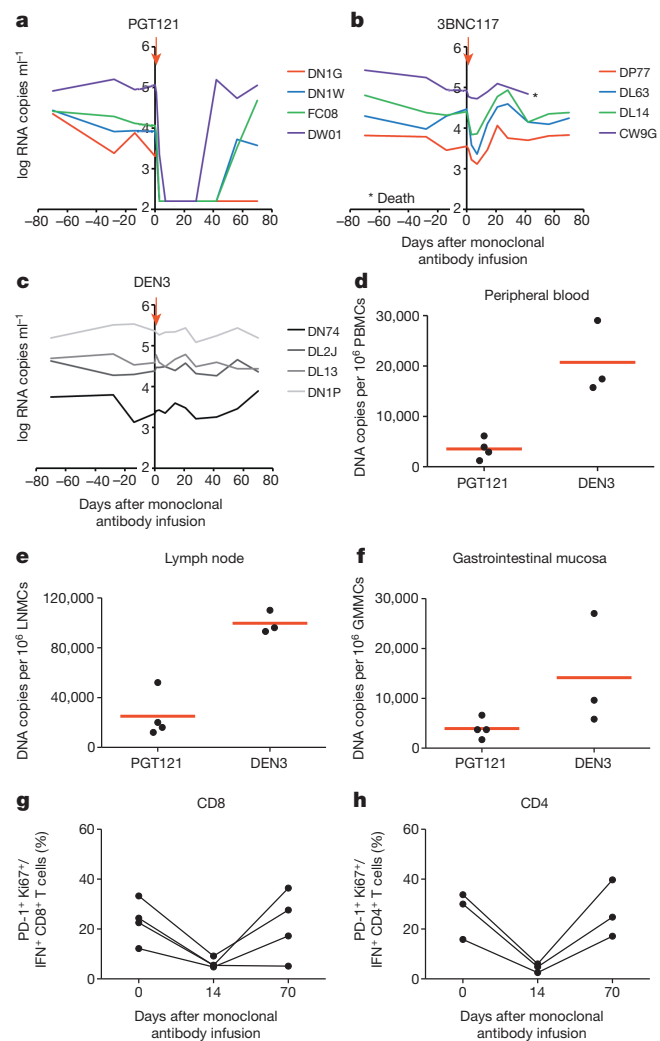
**Figure 2 | Therapeutic efficacy of the double PGT121 and 3BNC117 monoclonal antibody cocktail.** **a–c**, Plasma viral RNA (log copies ml<sup>-1</sup>) in rhesus monkeys chronically infected with SHIV-SF162P3 after a single infusion (arrows) of PGT121, 3BNC117 and b12 (**a**); PGT121 and 3BNC117 (**b**); or the control monoclonal antibody DEN3 (**c**). **d**, Plasma viral RNA in monkeys after a second infusion of PGT121 and 3BNC117 on day 105. **e, f**, PD-1<sup>+</sup>Ki67<sup>+</sup> expression on Gag-specific CD8<sup>+</sup> (**e**) and CD4<sup>+</sup> (**f**) T lymphocytes in the monkeys that received PGT121 and 3BNC117. **g, h**, Proviral DNA (copies per 10<sup>6</sup> cells) in lymph nodes (**g**) and gastrointestinal mucosa (**h**) before (day 105) and 14 days after (day 119) the second monoclonal antibody infusion with PGT121 and 3BNC117 in the four animals with detectable viraemia. GMMCs, gastrointestinal mucosa mononuclear cells; LNCMs, lymph node mononuclear cells.

stock was largely resistant to 3BNC117 (Extended Data Fig. 6), which raised the possibility that the observed therapeutic efficacy of the monoclonal antibody cocktail in the previous experiment may have been due to PGT121 alone. We therefore performed a single infusion of 10 mg kg<sup>-1</sup> PGT121 alone ( $n = 4$ ), 3BNC117 alone ( $n = 4$ ), or the control monoclonal antibody DEN3 ( $n = 4$ ) in 12 rhesus monkeys that were infected with SHIV-SF162P3 9 months before the monoclonal antibody infusion with chronic set-point viral loads of 3.3–5.4 log RNA copies ml<sup>-1</sup>. PGT121 alone resulted in rapid virological control to undetectable levels by day 7 in 4 of 4 animals, followed by viral rebound by day 42 to day 56 in 3 animals that again correlated with declines in serum PGT121 titres to undetectable levels (Fig. 3a, c and Extended Data Fig. 7;  $P = 0.02$  comparing viral loads on day 14 in PGT121-treated animals compared with controls). One animal exhibited long-term virological control (monkey DN1G; Fig. 3a). PGT121 alone

also reduced proviral DNA by sixfold in peripheral blood ( $P = 0.05$ ; Fig. 3d), fourfold in lymph nodes ( $P = 0.05$ ; Fig. 3e), and fourfold in gastrointestinal mucosa ( $P = 0.1$ ; Fig. 3f) as compared with the DEN3 control on day 14. Moreover, PGT121 alone resulted in threefold and fivefold reductions in the percentage of PD-1<sup>+</sup>Ki67<sup>+</sup> Gag-specific CD8<sup>+</sup> and CD4<sup>+</sup> T lymphocytes, respectively (Fig. 3g, h). In contrast, 3BNC117 alone, to which our SHIV-SF162P3 stock was relatively resistant, resulted in only a transient 0.2–1.1 log reduction of plasma viral loads (Fig. 3b), and one animal in this group (monkey CW9G) was euthanized for progressive clinical AIDS during this experiment.

### Kinetics of virological control

The kinetics of the initial decline of plasma viraemia after infusion of PGT121 or PGT121-containing monoclonal antibody cocktails was a median of 0.382 logs per day (inter-quartile range (IQR) 0.338–0.540). In contrast, the initial kinetics of decline of plasma viraemia after raltegravir-containing combination antiretroviral therapy in HIV-1-infected humans was a median of 0.264 logs per day (IQR 0.253–0.284)<sup>28</sup>



**Figure 3 | Therapeutic efficacy of the single monoclonal antibodies PGT121 and 3BNC117.** **a–c**, Plasma viral RNA (log copies ml<sup>-1</sup>) in rhesus monkeys chronically infected with SHIV-SF162P3 after a single infusion (arrows) of PGT121 (**a**), 3BNC117 (**b**), or the control monoclonal antibody DEN3 (**c**). **d–f**, Proviral DNA (copies per 10<sup>6</sup> cells) in peripheral blood mononuclear cells (PBMCs) (**d**), lymph nodes (**e**) and gastrointestinal mucosa (**f**) 14 days after the monoclonal antibody infusion in the animals that received PGT121 or DEN3. Red bars indicate mean values. Assays for one of the DEN3-treated controls failed. **g, h**, PD-1<sup>+</sup>Ki67<sup>+</sup> expression on Gag-specific CD8<sup>+</sup> (**g**) and CD4<sup>+</sup> (**h**) T lymphocytes in the monkeys that received PGT121.

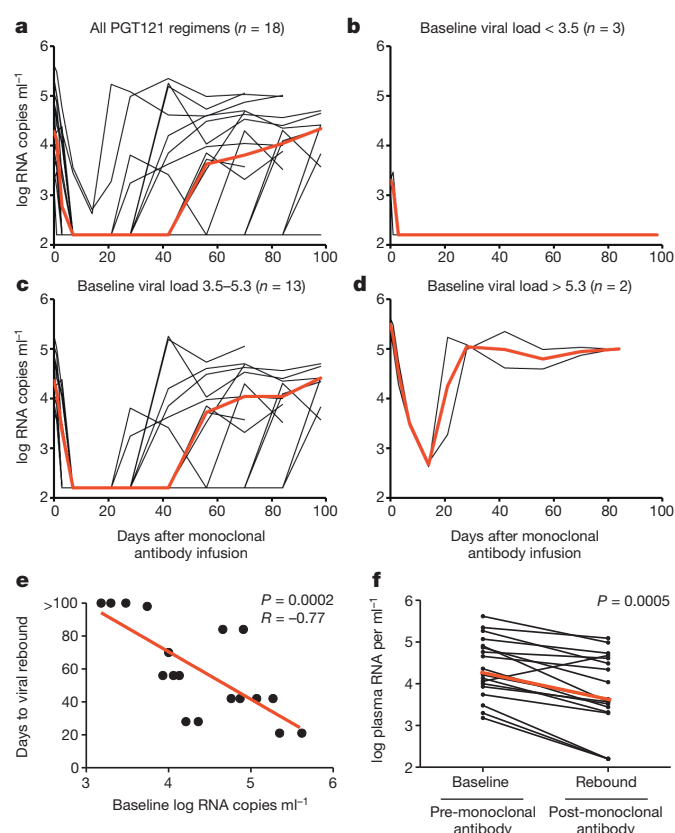
and after combination antiretroviral therapy in SIV-infected monkeys was a median of 0.229 logs per day (IQR 0.198–0.265) (J.B.W., unpublished data) (Extended Data Table 1). Although these reflect different models, the rapid virological control after monoclonal antibody administration in SHIV-infected rhesus monkeys is striking and consistent with a mechanism that involves direct elimination of free virus in plasma in addition to virus-infected cells in tissues. The rapid reduction of proviral DNA in peripheral blood by day 1 (Fig. 1e) suggests direct antibody-mediated cytotoxic effects on infected cells<sup>29</sup>, although we cannot exclude the possibility of an effect of CD4<sup>+</sup> T-cell trafficking.

## Summary and implications

Our studies demonstrate the profound therapeutic efficacy of PGT121 and PGT121-containing monoclonal antibody cocktails in chronically SHIV-SF162P3 infected rhesus monkeys. The therapeutic efficacy in the 18 animals that received PGT121 alone or as part of a cocktail (Fig. 4a) was dependent on baseline viral loads before monoclonal antibody administration. In the 17% of animals (3 of 18) with low baseline viral loads <3.5 log RNA copies ml<sup>-1</sup>, long-term control of viral replication was observed for the duration of the follow-up period (Fig. 4b), which included a substantial period of time after serum monoclonal antibody titres had declined to undetectable levels. These observations suggest that PGT121 may have converted animals with low baseline viraemia into ‘elite controllers’, although additional follow-up is required to assess the durability of this effect. These animals still had detectable, albeit reduced, proviral DNA in tissues (Figs 2g, h and 3e, f), and thus virus was not eradicated in these animals. In the 72% of animals (13 of 18) with intermediate baseline viral loads 3.5–5.3 log RNA copies ml<sup>-1</sup>, plasma viraemia was rapidly reduced to undetectable levels within 7 days but then rebounded after a median of 56 days when serum monoclonal antibody titres declined to undetectable levels <1 µg ml<sup>-1</sup> (Fig. 4c). In the 11% of animals (2 of 18) with high baseline viral loads >5.3 log RNA copies ml<sup>-1</sup>, incomplete control of plasma viraemia and rapid viral rebound was observed, indicating a therapeutic ceiling in this model (Fig. 4d). Taken together, baseline viral loads strongly correlated with the time to viral rebound ( $P = 0.0002$ , Spearman rank-correlation test; Fig. 4e).

We speculate that the therapeutic impact of these monoclonal antibodies reflected not only their direct antiviral activity but also their impact on host antiviral immune responses. After monoclonal antibody infusion, we observed modest increases in host virus-specific neutralizing antibody activity (Extended Data Fig. 2) as well as reduced activation and improved functionality of host virus-specific T-lymphocyte responses (Figs 1f–h, 2e, f and 3g, h). Consistent with these findings, median set-point viral loads after viral rebound were 0.61 log lower than median set-point viral loads at baseline before monoclonal antibody infusion ( $P = 0.0005$ , Wilcoxon matched pairs signed-rank test; Fig. 4f), with no evidence of reduced viral replicative capacity<sup>30</sup> (Extended Data Fig. 8). Moreover, 3 of 18 monkeys exhibited long-term virological control to undetectable levels (Fig. 4b). Defining the precise immunological mechanisms of improved virological control after monoclonal and polyclonal antibody administration<sup>31,32</sup> warrants further investigation.

Previous studies in humanized mice and humans showed that the earlier generation of neutralizing HIV-1-specific monoclonal antibodies was unable to control viraemia<sup>15–17</sup>. More recent studies in humanized mice have shown that combinations of three or five of the new generation of more potent monoclonal antibodies suppressed HIV-1 replication, although single monoclonal antibodies still rapidly selected for resistance<sup>18,19</sup>. In contrast to these previous studies, we observed that a single infusion of PGT121 in rhesus monkeys resulted in rapid virological control in both peripheral blood and tissues. It is possible that intrinsic differences between HIV-1 replication in mice and SHIV replication in monkeys may account for these differences. Another key difference is the functional immune system in rhesus monkeys as compared with immunosuppressed humanized mice, and the profound virological suppression without the development of



**Figure 4 | Therapeutic efficacy of PGT121 or PGT121-containing monoclonal antibody cocktails in chronically SHIV-infected rhesus monkeys.** **a–d**, Summary of the therapeutic effect of PGT121 alone or PGT121-containing monoclonal antibody cocktails in the 18 rhesus monkeys chronically infected with SHIV-SF162P3 (**a**), as well as in the subgroups of animals with baseline viral loads of <3.5 log RNA copies ml<sup>-1</sup> (**b**), 3.5–5.3 log RNA copies ml<sup>-1</sup> (**c**) and >5.3 log RNA copies ml<sup>-1</sup> (**d**). Red lines indicate median log viral loads. **e**, Correlation of baseline viral loads with times to viral rebound.  $P$  value reflects two-sided Spearman rank-correlation test. **f**, Comparison of set-point viral loads at baseline before monoclonal antibody administration and after viral rebound.  $P$  value reflects two-sided Wilcoxon matched pairs signed-rank test. Red line indicates median log viral loads.

resistance in the present study may reflect the functional host antibody effector activity and the intact antiviral cellular immune responses in rhesus monkeys. A caveat is that we were unable to quantify the intrinsic ability of SHIV-SF162P3 to escape from PGT121 *in vivo*, although previous studies have documented the ability of SHIV-SF162P3 and SHIV-SF162P4 to escape from autologous antibody responses in other settings<sup>33,34</sup>. N332A-mutated SHIV-SF162P3 exhibited only partial escape from PGT121 *in vitro* but complete escape from other N332-dependent monoclonal antibodies including PGT124 and PGT128, suggesting a high bar to resistance (Extended Data Fig. 9).

## Concluding remarks

Our data demonstrate profound therapeutic efficacy of broad and potent HIV-1-specific monoclonal antibodies in rhesus monkeys chronically infected with SHIV-SF162P3. Our SHIV-SF162P3 stock is highly pathogenic, as evidenced by moderate to high chronic set-point viral loads and AIDS-related deaths in two animals during the course of these experiments (Figs 2c and 3b). Moreover, in a separate study<sup>35</sup>, our SHIV-SF162P3 stock led to AIDS-related mortality in 5 of 12 (42%) rhesus monkeys by 1 year after infection, which is comparable to the reported pathogenicity of SIVmac251 (refs 36, 37) and greater than the reported mortality of SHIV-AD8 (refs 38, 39). Nevertheless, given the differences between SHIV-infected rhesus



monkeys and HIV-1-infected humans, clinical trials will be required to establish the therapeutic efficacy of potent neutralizing HIV-1-specific monoclonal antibodies in humans. Although multiple monoclonal antibodies targeting different epitopes will probably prove to be superior, our data suggest that monotherapy with the most potent and broad monoclonal antibodies also warrants clinical evaluation. Moreover, the ability of these monoclonal antibodies to reduce proviral DNA in tissues suggests that these monoclonal antibodies should also be evaluated in the context of viral eradication strategies.

## METHODS SUMMARY

Thirty-four Indian origin, young adult, male and female rhesus monkeys (*Macaca mulatta*) that were infected by the intrarectal route with SHIV-SF162P3 and followed 9 months before monoclonal antibody administration were used for these studies. Cocktails of monoclonal antibodies (PGT121, 3BNC117, b12) or single monoclonal antibodies (PGT121, 3BNC117) were administered once or twice at 10 mg kg<sup>-1</sup> for each monoclonal antibody by the intravenous route. DEN3 was used as an isotype-matched monoclonal antibody control. Monkeys were bled up to three times per week for evaluation of plasma viral loads. Lymph node and gastrointestinal mucosal biopsies were processed as single cell suspensions<sup>27</sup> for proviral DNA assays as previously described<sup>26</sup>. SIV Gag-specific cellular immune responses were assessed by multiparameter intracellular cytokine staining (ICS) assays<sup>23,24</sup> and functional virus suppression assays<sup>25</sup> essentially as described. HIV-1-specific neutralizing antibody responses were assessed by TZM-bl luciferase-based neutralization assays<sup>22</sup>. PGT121, 3BNC117 and b12 concentrations were determined by monoclonal-antibody-specific neutralizing antibody assays. Virus sequencing and analyses of breakthrough viruses were performed essentially as described<sup>18,21</sup>. Statistical analyses involved Mann–Whitney *U*-tests for comparing independent sets of virological and immunologic parameters, Wilcoxon matched pairs signed-rank tests for paired analyses, and Spearman rank-correlation tests for correlation analyses.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 July; accepted 8 October 2013.

Published online 30 October 2013.

- Burton, D. R., Poignard, P., Stanfield, R. L. & Wilson, I. A. Broadly neutralizing antibodies present new prospects to counter highly antigenically diverse viruses. *Science* **337**, 183–186 (2012).
- Klein, F. *et al.* Antibodies in HIV-1 vaccine development and therapy. *Science* **341**, 1199–1204 (2013).
- Wu, X. *et al.* Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* **329**, 856–861 (2010).
- Zhou, T. *et al.* Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science* **329**, 811–817 (2010).
- Scheid, J. F. *et al.* Broad diversity of neutralizing antibodies isolated from memory B cells in HIV-infected individuals. *Nature* **458**, 636–640 (2009).
- Scheid, J. F. *et al.* Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* **333**, 1633–1637 (2011).
- Diskin, R. *et al.* Increasing the potency and breadth of an HIV antibody by using structure-based rational design. *Science* **334**, 1289–1293 (2011).
- Walker, L. M. *et al.* Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science* **326**, 285–289 (2009).
- McLellan, J. S. *et al.* Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature* **480**, 336–343 (2011).
- Walker, L. M. *et al.* Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* **477**, 466–470 (2011).
- Julien, J. P. *et al.* Broadly neutralizing antibody PGT121 allosterically modulates CD4 binding via recognition of the HIV-1 gp120 V3 base and multiple surrounding glycans. *PLoS Pathog.* **9**, e1003342 (2013).
- Mouquet, H. *et al.* Complex-type N-glycan recognition by potent broadly neutralizing HIV antibodies. *Proc. Natl Acad. Sci. USA* **109**, E3268–E3277 (2012).
- Kong, L. *et al.* Supersite of immune vulnerability on the glycosylated face of HIV-1 envelope glycoprotein gp120. *Nature Struct. Mol. Biol.* **20**, 796–803 (2013).
- Huang, J. *et al.* Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature* **491**, 406–412 (2012).
- Poignard, P. *et al.* Neutralizing antibodies have limited effects on the control of established HIV-1 infection *in vivo*. *Immunity* **10**, 431–438 (1999).
- Trkola, A. *et al.* Delay of HIV-1 rebound after cessation of antiretroviral therapy through passive transfer of human neutralizing antibodies. *Nature Med.* **11**, 615–622 (2005).
- Mehandru, S. *et al.* Adjunctive passive immunotherapy in human immunodeficiency virus type 1-infected individuals treated with antiviral therapy during acute and early infection. *J. Virol.* **81**, 11016–11031 (2007).
- Klein, F. *et al.* HIV therapy by a combination of broadly neutralizing antibodies in humanized mice. *Nature* **492**, 118–122 (2012).
- Diskin, R. *et al.* Restricting HIV-1 pathways for escape using rationally designed anti-HIV-1 antibodies. *J. Exp. Med.* **210**, 1235–1249 (2013).
- Burton, D. R. *et al.* Efficient neutralization of primary isolates of HIV-1 by a recombinant human monoclonal antibody. *Science* **266**, 1024–1027 (1994).
- West, A. P. Jr *et al.* Computational analysis of anti-HIV-1 antibody neutralization panel data to identify potential functional epitope residues. *Proc. Natl Acad. Sci. USA* **110**, 10598–10603 (2013).
- Montefiori, D. *Evaluating Neutralizing Antibodies Against HIV, SIV and SHIV in Luciferase Reporter Gene Assays. Current Protocols Immunol* (John Wiley & Sons, 2004).
- Pitcher, C. J. *et al.* Development and homeostasis of T cell memory in rhesus macaque. *J. Immunol.* **168**, 29–43 (2002).
- Liu, J. *et al.* Magnitude and phenotype of cellular immune responses elicited by recombinant adenovirus vectors and heterologous prime-boost regimens in rhesus monkeys. *J. Virol.* **82**, 4844–4852 (2008).
- Stephenson, K. E., Li, H., Walker, B. D., Michael, N. L. & Barouch, D. H. Gag-specific cellular immunity determines *in vitro* viral inhibition and *in vivo* virologic control following simian immunodeficiency virus challenges of vaccinated rhesus monkeys. *J. Virol.* **86**, 9583–9589 (2012).
- Whitney, J. B. *et al.* T-cell vaccination reduces simian immunodeficiency virus levels in semen. *J. Virol.* **83**, 10840–10843 (2009).
- Li, H. *et al.* Durable mucosal simian immunodeficiency virus-specific effector memory T lymphocyte responses elicited by recombinant adenovirus vectors in rhesus monkeys. *J. Virol.* **85**, 11007–11015 (2011).
- Andrade, A. *et al.* Three distinct phases of HIV-1 RNA decay in treatment-naïve patients receiving raltegravir-based antiretroviral therapy: ACTG A5248. *J. Infect. Dis.* **208**, 884–891 (2013).
- Horwitz, J. A. *et al.* HIV-1 suppression and durable control by combining single broadly neutralizing antibodies and antiretroviral drugs in humanized mice. *Proc. Natl Acad. Sci. USA* **110**, 16538–16543 (2013).
- Cornall, A. *et al.* A novel, rapid method to detect infectious HIV-1 from plasma of persons infected with HIV-1. *J. Virol. Methods* **165**, 90–96 (2010).
- Jaworski, J. P. *et al.* Neutralizing polyclonal IgG present during acute infection prevents rapid disease onset in simian-human immunodeficiency virus SHIVSF162P3-infected infant rhesus macaques. *J. Virol.* **87**, 10447–10459 (2013).
- Ng, C. T. *et al.* Passive neutralizing antibody controls SHIV viremia and enhances B cell responses in infant macaques. *Nature Med.* **16**, 1117–1119 (2010).
- Jayaraman, P. *et al.* Evidence for persistent, occult infection in neonatal macaques following perinatal transmission of simian-human immunodeficiency virus SF162P3. *J. Virol.* **81**, 822–834 (2007).
- Kraft, Z. *et al.* Macaques infected with a CCR5-tropic simian/human immunodeficiency virus (SHIV) develop broadly reactive anti-HIV neutralizing antibodies. *J. Virol.* **81**, 6402–6411 (2007).
- Barouch, D. H. *et al.* Protective efficacy of a global HIV-1 mosaic vaccine against heterologous SHIV challenges in rhesus monkeys. *Cell* **155**, 531–539 (2013).
- Barouch, D. H. *et al.* Vaccine protection against acquisition of neutralization-resistant SIV challenges in rhesus monkeys. *Nature* **482**, 89–93 (2012).
- Liu, J. *et al.* Immune control of an SIV challenge by a T-cell-based vaccine in rhesus monkeys. *Nature* **457**, 87–91 (2009).
- Nishimura, Y. *et al.* Generation of the pathogenic R5-tropic simian/human immunodeficiency virus SHIVAD8 by serial passaging in rhesus macaques. *J. Virol.* **84**, 4769–4781 (2010).
- Gautam, R. *et al.* Pathogenicity and mucosal transmissibility of the R5-tropic simian/human immunodeficiency virus SHIV(AD8) in rhesus macaques: implications for use in vaccine studies. *J. Virol.* **86**, 8516–8526 (2012).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank A. Brinkman, M. Ferguson, C. Gittens, R. Gelezianus, R. Hamel, K. Kelly, J. Kramer, A. McNally, D. Montefiori, L. Nogueira, L. Parenteau, M. Pensiero, L. Peter, M. Shetty, D. Sok, K. Stanley, F. Stephens, W. Wagner, B. Walker, A. West and J. Yalley-Ogunro for advice, assistance and reagents. The SIVmac239 Gag peptide pool was obtained from the NIH AIDS Research and Reference Reagent Program. We acknowledge support from the National Institutes of Health (AI055332, AI060354, AI078526, AI084794, AI095985, AI096040, AI10063, AI100148, AI100663); the Bill and Melinda Gates Foundation (OPP1033091, OPP1033115, OPP1040741, OPP1040753); the Ragon Institute of MGH, MIT, and Harvard; the Lundbeck Foundation; and the Stavros Niarchos Foundation. M.C.N. is a Howard Hughes Medical Institute investigator. M.C.N. and D.R.B. are co-inventors on patents covering the monoclonal antibodies used in the present study.

**Author Contributions** D.H.B., M.C.N. and D.R.B. designed the studies. J.B.W., B.M., F.K., T.Y.O., H.-W.C., S.S. and P.P. led the virological assays. B.M., J.L., K.E.S., M.S.S., K.M.S., E.N.B., C.C., J.Y.S., S.B. and J.R.P. led the immunological assays. K.S., S.G. and A.K.C. led the kinetic analyses. J.B.W., J.P.N., M.B., M.G.L. and W.R. led the monoclonal antibody infusions and clinical care of the rhesus monkeys. D.H.B. led the studies and wrote the paper with all co-authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.H.B. ([dbarouch@bidmc.harvard.edu](mailto:dbarouch@bidmc.harvard.edu)).

## METHODS

**Animals and study design.** Thirty-four Indian-origin, outbred, young adult, male and female, specific pathogen-free rhesus monkeys (*Macaca mulatta*) that did not express the class I alleles *Mamu-A\*01*, *Mamu-B\*08* and *Mamu-B\*17* associated with spontaneous virological control were housed at New England Primate Research Center, Bioqual, or Alpha Genesis. Groups were balanced for susceptible and resistant TRIM5 $\alpha$  alleles. Groups of 4–5 monkeys provided statistical power to detect large differences in viral loads, and animals were randomly allocated to balance baseline viral loads. Animals were infected by the intrarectal route with our rhesus-derived SHIV-SF162P3 challenge stock 9 months before monoclonal antibody administration. PGT121, b12 and DEN3 monoclonal antibodies were generated as previously described<sup>10</sup> and were expressed in Chinese hamster ovary (CHO-K1) cells and purified by protein A affinity chromatography. 3BNC117 was manufactured by Celldex Therapeutics in CHO cells and purified by chromatography and sterile filtration. All the monoclonal antibody preparations were endotoxin free. Cocktails of monoclonal antibodies or single monoclonal antibodies were administered to monkeys once or twice by the intravenous route at a dose of 10 mg kg<sup>-1</sup> for each monoclonal antibody. Monkeys were bled up to three times per week for viral loads. Immunological and virological data were generated blinded. All animal studies were approved by the appropriate Institutional Animal Care and Use Committee (IACUC).

**Cellular immune assays.** SIV Gag-specific cellular immune responses were assessed by multiparameter intracellular cytokine staining (ICS) assays essentially as described<sup>23,24</sup>. 12-colour ICS assays were performed with the Aqua green-fluorescent reactive dye (Invitrogen, L23101) and predetermined titres of monoclonal antibodies (Becton Dickinson) against CD3 (SP34; Alexa Fluor 700), CD4 (OKT4; BV711, Biolegend), CD8 (SK1; allophycocyanin-cyanine 7 (APC-Cy7)), CD28 (L293; BV610), CD95 (DX2; allophycocyanin), CD69 (TP1.55.3; phycoerythrin-Texas red (energy-coupled dye; ECD); Beckman Coulter), gamma interferon (IFN- $\gamma$ ) (B27; phycoerythrin-cyanine 7 (PE-Cy7)), Ki67 (B56; fluorescein isothiocyanate (FITC)), CCR5 (3A9; phycoerythrin), CCR7 (3D12; Pacific Blue) and PD-1 (EH21.1; peridinin chlorophyll-A-cyanine 5.5 (PerCP-Cy5.5)). IFN- $\gamma$  backgrounds were consistently <0.01% in PBMCs and LNMCs and <0.05% in colorectal biopsy specimens. Virus suppression assays were performed by co-culturing purified CD8<sup>+</sup> T lymphocytes with CD8-depleted PBMCs and monitoring p27 levels for 7–14 days essentially as described<sup>25</sup>.

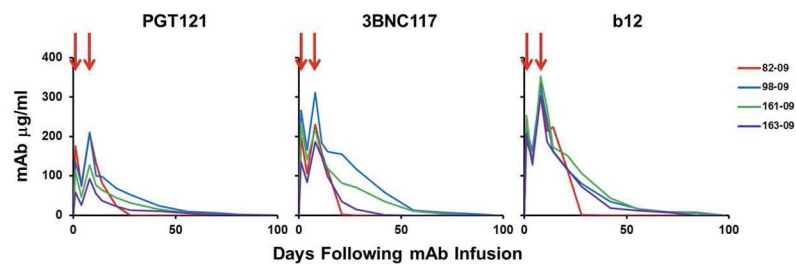
**Neutralizing antibody assays.** HIV-1-specific neutralizing antibody responses against primary infectious stocks of SHIV-SF162P3 and SHIV-SF162P4 were assessed by TZM-bl luciferase-based neutralization assays<sup>22</sup>. PGT121 titres were determined by X2088\_c9 and ZM247v1 (Rev-) pseudovirus neutralization, 3BNC117 titres were determined by 6041.v3.c23 and Q461.ez pseudovirus neutralization,

and b12 titres were determined by Du422.1.N332A pseudovirus neutralization and B2.1 ELISA.

**Proviral DNA assay.** Proviral DNA was quantified as previously reported<sup>26</sup>. Lymph node and gastrointestinal mucosal biopsies were processed as single cell suspensions essentially as previously described<sup>27</sup>. Tissue-specific total cellular DNA was isolated from  $5 \times 10^6$  cells using a QIAamp DNA Blood Mini kit (Qiagen). The absolute quantification of viral DNA in each sample was determined by qPCR using primers specific to a conserved region SIVmac239. All samples were directly compared to a linear virus standard and the simultaneous amplification of a fragment of human *GAPDH* gene. The sensitivity of linear standards was compared against the 3D8 cell line as a reference standard as described<sup>26</sup>. All PCR assays were performed with 100 and 200 ng of sample DNA.

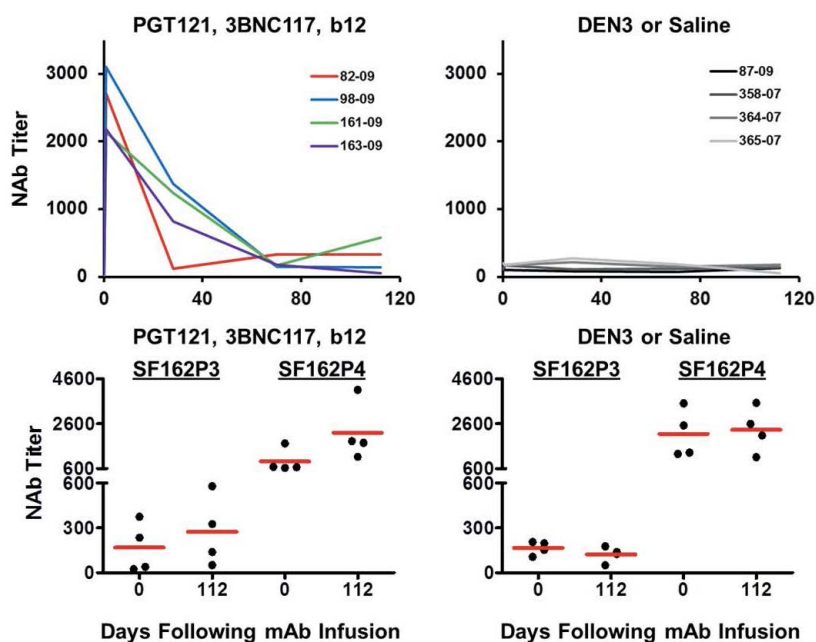
**Virus sequencing.** Virus sequencing of breakthrough virus was performed essentially as described<sup>18</sup>. Plasma samples of 1 ml were centrifuged for 30 min at 20,000g and the lowest fraction was subjected to RNA purification (QiaAmp MinElute Virus Spin kit; Qiagen). Random hexamers (Roche) or SHIV-SF162P3-specific (5'-AAGAGCTCCTCCAGACAGTGAG-3' or 5'-TAGAGCCTGGAAGCATCCAGGAAGTCAGCCTA-3') primers were used for cDNA synthesis with SuperScript III Reverse Transcriptase (Invitrogen). SHIV envelope sequences were amplified by a double-nested PCR approach using the Expand High Fidelity PCR System (Roche). First round primers for gp120 were 5'-AAGAGCTCCTCCAGACAGTGAG-3' and 5'-ATGAGTTTCCAGAGCAACCC-3' and for gp160 were 5'-AAGAGCTCCTCCAGACAGTGAG-3' and 5'-CAAGCCCTTGTCTAATCCTCC-3'. Second round primers for gp120 were 5'-GAAAGAGCAGAAGACAGTGGC-3' and 5'-ATTGTCTGGCCTGTACCGTC-3' and for gp160 were 5'-GAAAGAGCAGAAGACAGTGGC-3' and 5'-ATGGAAATAGCTCCA CCCATC-3'. After second round PCR, all products were spiked with 0.5  $\mu$ l Taq polymerase and incubated for 15 min at 72 °C. Amplicons were excised from a gel and purified after cloning into the pCR4-TOPO vector (Invitrogen) and expansion in One Shot TOP10 cells at 30 °C. Single colonies were sequenced using M13F/M13R primers as well as primers annealing to the envelope sequence. A consensus sequence of each clone was derived using Geneious Pro software (Biomatters), and sequence analysis was performed using Geneious Pro and antibody database software<sup>21</sup>.

**Statistical analyses.** Analyses of independent virological and immunological data were performed by two-tailed Mann–Whitney *U*-tests. Analyses of paired data sets were performed by two-tailed Wilcoxon matched pairs signed-rank tests. Correlations were evaluated by Spearman rank-correlation tests. *P* values less than 0.05 were considered significant. Statistical analyses were performed using GraphPad Prism. Exponential decay rates of plasma viral loads were calculated using standard ordinary least squares regression on log<sub>10</sub> (viral load) measurements versus time (days).

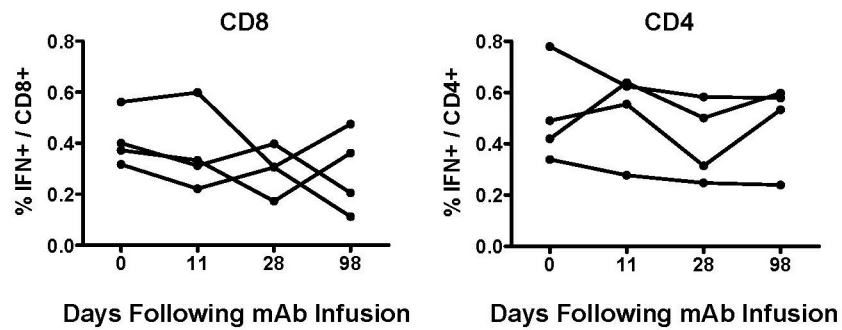


**Extended Data Figure 1 | Monoclonal antibody titres after administration of the triple PGT121, 3BNC117 and b12 monoclonal antibody cocktail.** PGT121, 3BNC117 and b12 titres in the monkeys described in Fig. 1 after infusion of the triple monoclonal antibody cocktail (arrows).

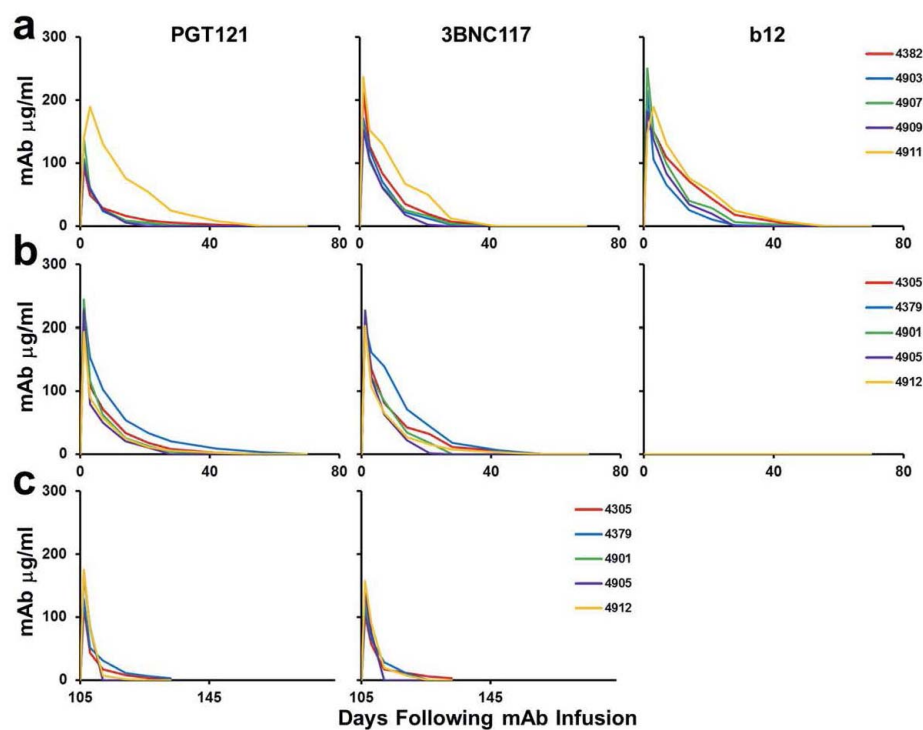




**Extended Data Figure 2 | Neutralizing antibody titres after administration of the triple PGT121, 3BNC117 and b12 monoclonal antibody cocktail.** SHIV-SF162P3 and SHIV-SF162P4 serum neutralizing antibody ID<sub>50</sub> titres in the monkeys described in Fig. 1 after infusion of the triple monoclonal antibody cocktail (left) or saline or DEN3 (right).



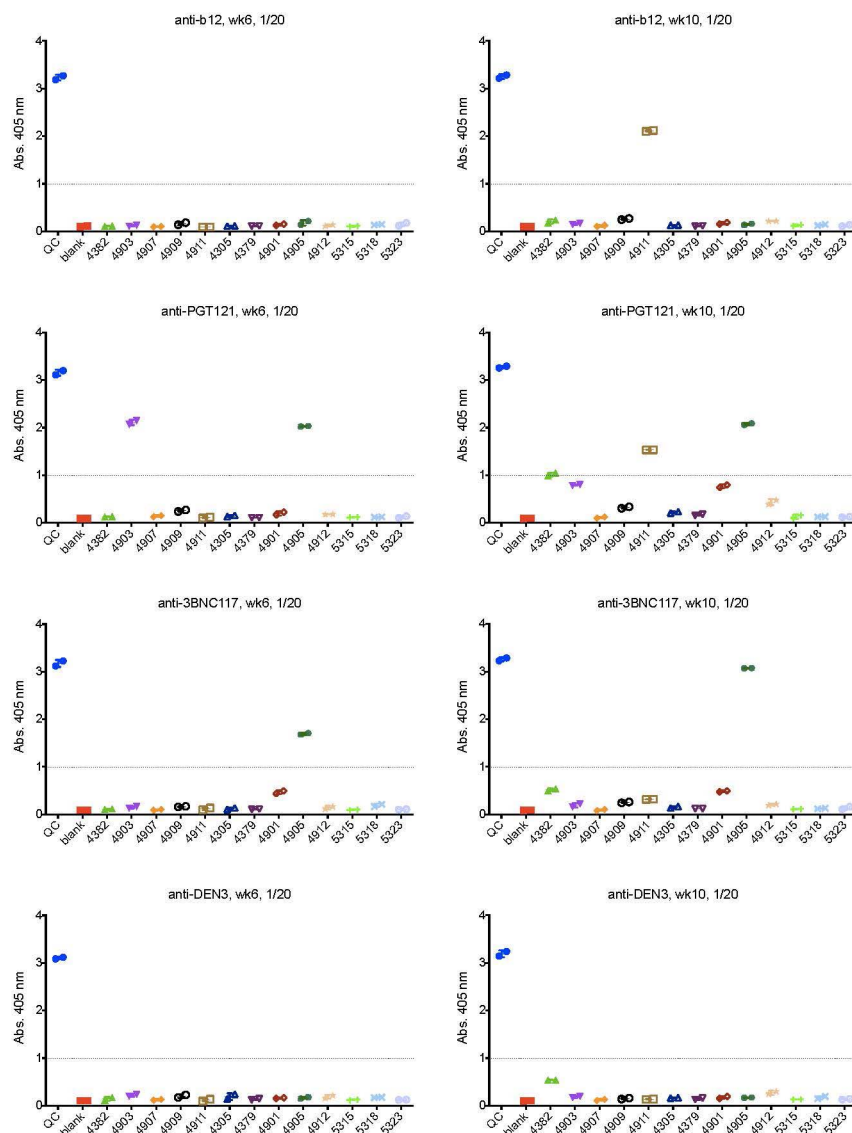
**Extended Data Figure 3** | Gag-specific T lymphocyte responses after administration of the triple PGT121, 3BNC117 and b12 monoclonal antibody cocktail. Gag-specific IFN- $\gamma$ <sup>+</sup> CD8<sup>+</sup> (left) and CD4<sup>+</sup> (right) T-lymphocyte responses by multiparameter intracellular cytokine staining assays in the monkeys described in Fig. 1 after infusion of the triple monoclonal antibody cocktail.



**Extended Data Figure 4 | Monoclonal antibody titres after administration of the double PGT121 and 3BNC117 monoclonal antibody cocktail.**  
a–c, PGT121, 3BNC117 and b12 titres in the monkeys described in Fig. 2 that

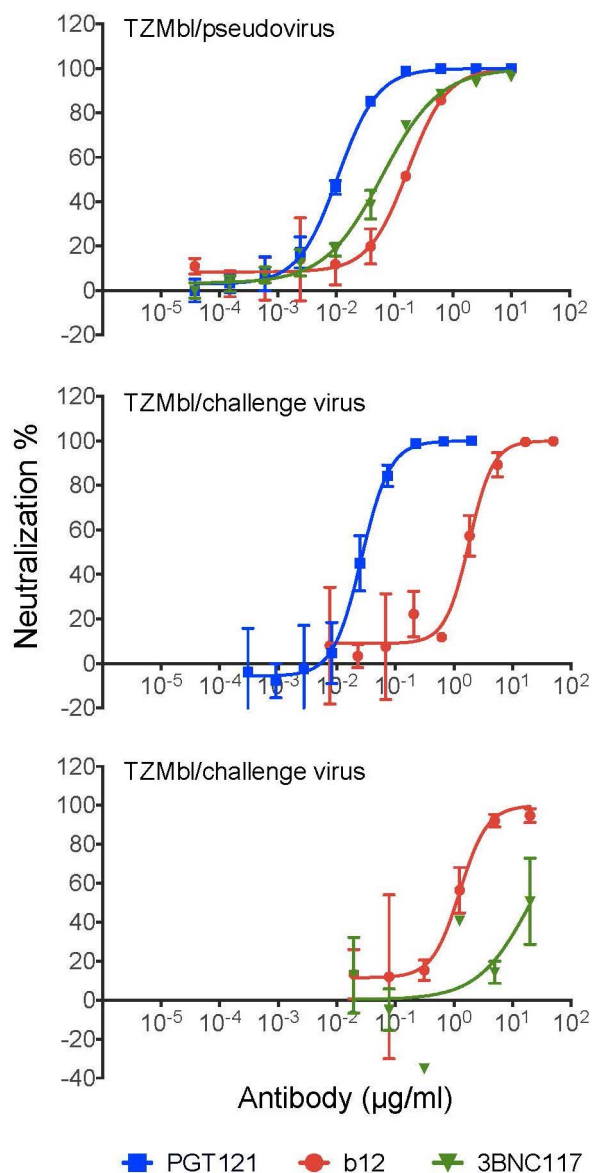
received PGT121, 3BNC117 and b12 (a); PGT121 and 3BNC117 (b); or the second infusion of PGT121 and 3BNC117 (c).



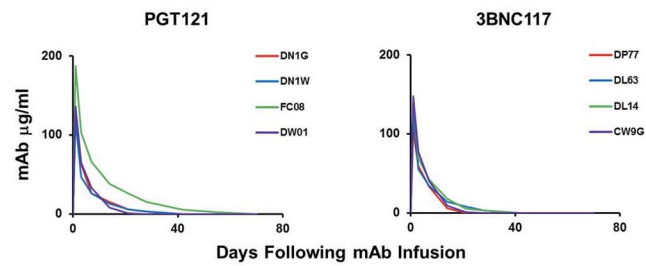


**Extended Data Figure 5 | Monkey anti-human antibody titres after monoclonal antibody administration.** ELISAs assessing anti-b12,

anti-PGT121, anti-3BNC117 and anti-DEN3 antibodies at week 6 and week 10 after monoclonal antibody infusion in the monkeys described in Fig. 2.

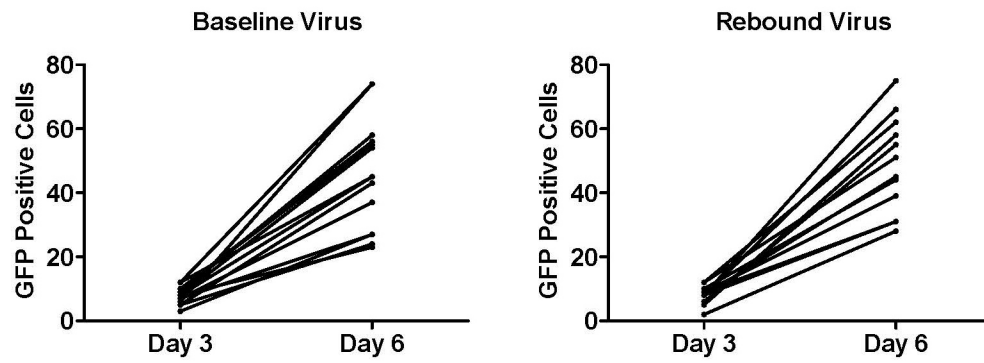


**Extended Data Figure 6 | Neutralization sensitivity of SHIV-SF162P3 pseudovirus and our SHIV-SF162P3 challenge stock.** TZM-bl neutralization assays of PGT121, 3BNC117 and b12 against the SHIV-SF162P3 pseudovirus (top) and against the SHIV-SF162P3 challenge stock (middle, bottom). Note sensitivity of our SHIV-SF162P3 challenge stock to PGT121 but relative resistance to 3BNC117.

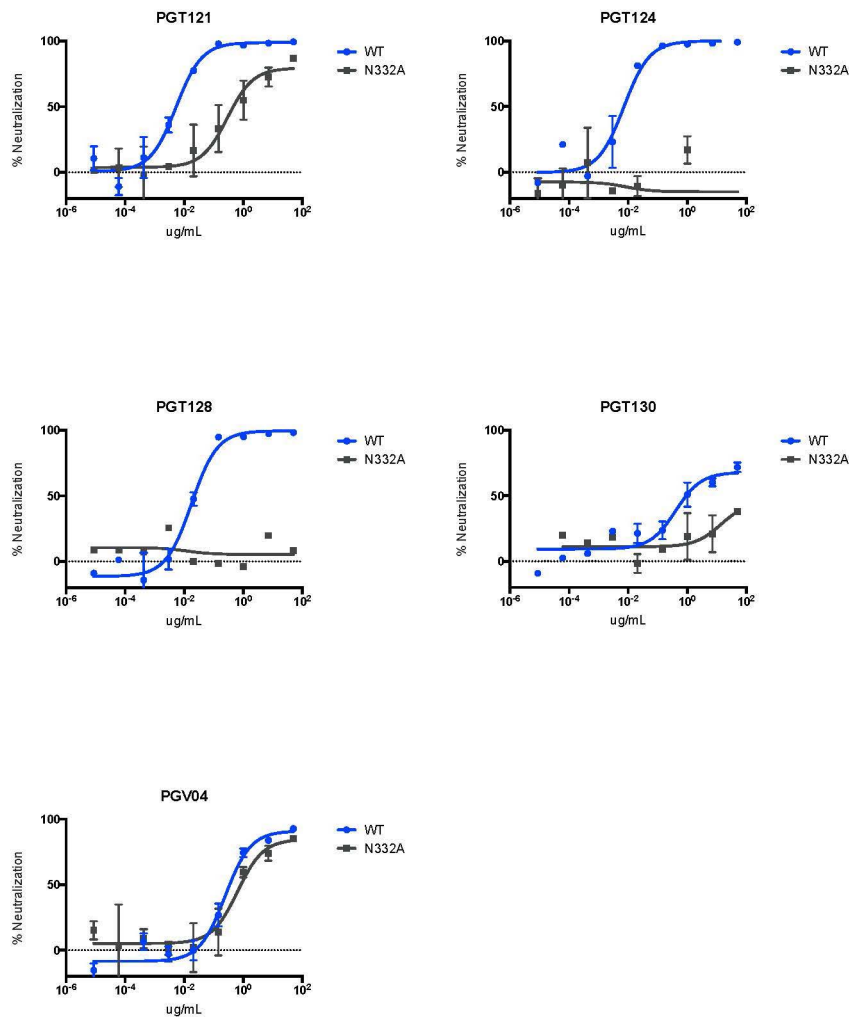


**Extended Data Figure 7 | Monoclonal antibody titres after administration of the single monoclonal antibodies PGT121 and 3BNC117.** PGT121 and 3BNC117 titres in the monkeys described in Fig. 3 after infusion of the single monoclonal antibodies.





**Extended Data Figure 8 | Virus replicative capacity at baseline and following virus rebound.** Numbers of GFP-positive infected GHOST indicator cells per well after 3 and 6 days of culture with baseline or rebound SHIV-SF162P3 virus.



**Extended Data Figure 9 | Monoclonal antibody sensitivity to N332A mutated SHIV-SF162P3.** TZM-bl neutralization assays of PGT121, PGT124, PGT128, PGT130 and PGV04 against SHIV-SF162P3 containing the N332A

mutation. Note 100-fold reduced sensitivity to PGT121 but more profound escape from PGT124 and PGT128.

**Extended Data Table 1 | Viral decay kinetics**

Therapy	$r$ (logs/day) <sup>†</sup> Median (IQR)	Median $t_{1/2}$ (days) <sup>#</sup>	Fold decrease in viral load over a 7-day period
DTG + TNF/FTC (SIV/monkeys)	0.229 (0.198-0.265) <sup>*</sup>	1.31	40
EFV + 2 NRTI (HIV/humans)	0.294 (0.273-0.334)	1.02	112
RAL + TDF/FTC (HIV/humans)	0.264 (0.253-0.284)	1.15	70
PGT121 mAb (SHIV/monkeys)	0.382 (0.338-0.540)	0.78	468

Kinetics of decline of plasma viraemia after antiretroviral therapy in rhesus monkeys and humans and after monoclonal antibody administration in rhesus monkeys. DTG, dolutegravir; EFV, efavirenz; FTC, emtricitabine; NRTI, nucleoside reverse transcriptase inhibitor; RAL, raltegravir; TDF, tenofovir disoproxil fumarate; TNF, tenofovir.

\* Computed using viral load measurements at day 0 and day 12 (J.B.W., unpublished).

† In the case of EFV and RAL therapies, decline rates  $r$  correspond to the 'first phase' of viral decline. Values reported in ref. 28 have been converted to logs/day (base 10). Note that the specific rate of decline due to RAL is slower than that due to EFV. The rapid viral decline in RAL compared to EFV is due to a longer duration in the first phase and a slower transition into the second phase, where viral decline rates are lower.

#The half-life,  $t_{1/2} = \ln(2)/r \cdot \ln(10)$ .



# RNA catalyses nuclear pre-mRNA splicing

Sebastian M. Fica<sup>1,2\*</sup>, Nicole Tuttle<sup>3\*</sup>, Thaddeus Novak<sup>4</sup>, Nan-Sheng Li<sup>4</sup>, Jun Lu<sup>3</sup>, Prakash Koodathingal<sup>2</sup>, Qing Dai<sup>3</sup>, Jonathan P. Staley<sup>2</sup> & Joseph A. Piccirilli<sup>3,4</sup>

**In nuclear pre-messenger RNA splicing, introns are excised by the spliceosome, a dynamic machine composed of both proteins and small nuclear RNAs (snRNAs). Over thirty years ago, after the discovery of self-splicing group II intron RNAs, the snRNAs were proposed to catalyse splicing. However, no definitive evidence for a role of either RNA or protein in catalysis by the spliceosome has been reported so far. By using metal rescue strategies in spliceosomes from budding yeast, here we show that the U6 snRNA catalyses both of the two splicing reactions by positioning divalent metals that stabilize the leaving groups during each reaction. Notably, all of the U6 catalytic metal ligands we identified correspond to the ligands observed to position catalytic, divalent metals in crystal structures of a group II intron RNA. These findings indicate that group II introns and the spliceosome share common catalytic mechanisms and probably common evolutionary origins. Our results demonstrate that RNA mediates catalysis within the spliceosome.**

Nuclear pre-mRNA splicing (Fig. 1a) is a crucial determinant of the export, translation, stability and diversity of eukaryotic messages<sup>1</sup>, but the spliceosome is the only major cellular machinery<sup>2</sup> required for gene expression for which the catalytic components remain undefined. Nevertheless, for three decades, there has been widespread speculation that nuclear pre-mRNA splicing is catalysed by RNA.

This speculation arose from the discovery of self-splicing RNAs, the identification of snRNA components of the spliceosome, and the finding that pre-mRNA introns and group II introns both splice through an intermediate having a lariat structure<sup>3,4</sup> (Fig. 1a). Since then, genetic, biochemical and NMR data have shown that the snRNAs share functional and structural similarity with the catalytic core of group II introns<sup>5–10</sup>. Similarly to the catalytic domain V of group II introns, U2/U6 helix Ib and the intramolecular stem-loop (ISL) of U6 adopt a secondary structure having a conserved bulge and AGC triad sensitive to phosphorothioate substitutions and important for both steps of splicing<sup>5–8,10–16</sup> (Fig. 1b, c). Extending the parallel, a recent crystal structure of a central splicing factor, Prp8, revealed domains similar to those found in cofactors of group II introns<sup>17</sup>.

Consistent with a catalytic role for RNA in the spliceosome, in the absence of spliceosomal proteins, U2 and U6 can base-pair and fold *in vitro* into a structure that catalyses reactions similar to the two steps of pre-mRNA splicing<sup>18,19</sup>, although the relevance of such protein-free, minimal model systems for understanding spliceosomal catalysis has been questioned (ref. 20 and compare with ref. 21).

Whether or not through RNA, the catalytic centre of the spliceosome, like that of group II introns, was proposed to catalyse the two phosphotransesterifications of splicing by positioning two catalytic metals<sup>22</sup>. In this two-metal mechanism<sup>22</sup>, one metal would stabilize the nucleophile and the second metal would stabilize the leaving group (Fig. 1a). Indeed, in human spliceosomes, as well as group II introns<sup>23</sup>, divalent metals stabilize the leaving group during each step of splicing<sup>24,25</sup>.

Intriguingly, recent crystal structures of a group II intron have revealed that domain V utilizes five non-bridging phosphate oxygens to coordinate two metals 3.9 Å apart<sup>26,27</sup>—ideally positioned to effect catalysis by the two-metal mechanism<sup>22</sup>. By analogy, the snRNAs have

been suggested to similarly position metals, consistent with early phosphorothioate substitution studies in U6 (refs 12, 13). However, only residue U80, situated in the U6 ISL, has been shown to interact with a metal<sup>14,28</sup> and it has remained unclear whether U80 positions a structural or a catalytic metal. Thus, despite work highlighting similarities between self-splicing RNA and the snRNAs, there is still no direct evidence that the snRNAs mediate splicing catalysis.

Definitive evidence for a direct role for metals coordinated by the RNA in the catalysis of self-splicing group I introns has come from metal rescue strategies<sup>29–31</sup>. These approaches, validated by subsequent structural studies<sup>31,32</sup>, enabled the direct linkage of metal ligands in the ribozyme to the splice sites. Application of such strategies in an investigation of pre-mRNA splicing has been hindered by proofreading and discard mechanisms that compete with catalysis during both steps of splicing<sup>28,33–35</sup>. Here by disabling such proofreading, we implemented metal rescue strategies in the fully assembled spliceosome and identified the direct effectors of splicing catalysis. Our results provide definitive evidence that snRNAs interact with the splice sites directly through catalytic metals during both chemical steps of splicing, establishing that the spliceosome utilizes RNA to catalyse splicing.

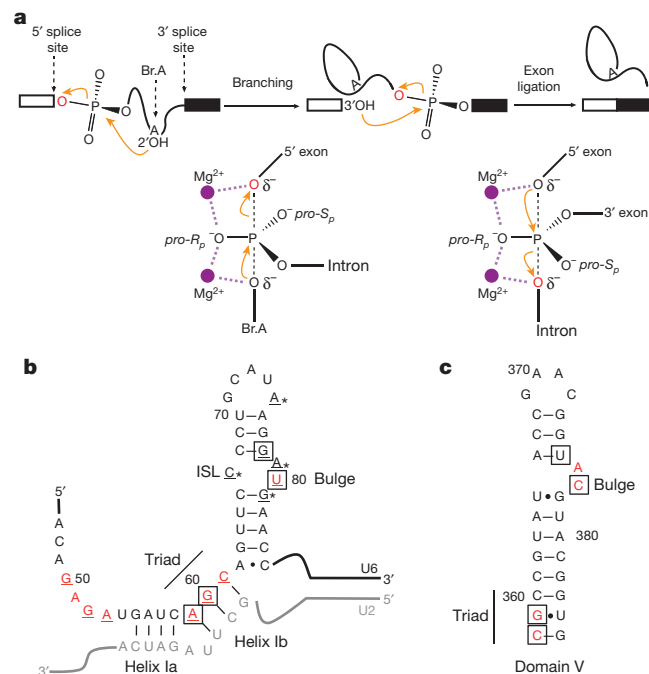
## U6 binds metals during both splicing steps

To identify components that mediate metal catalysis in the spliceosome, we used metal rescue approaches<sup>31</sup> (Fig. 2a) to find metal ligands that function at the catalytic stage and subsequently test these ligands for a direct role in positioning catalytic metals (Fig. 1a). The similarities between U6 and the catalytic domain V of group II introns suggest U6 as the best candidate for providing metal ligands that function at the catalytic stage (Fig. 1b, c). Indeed, early sulphur substitution studies implicated eight oxygens in the phosphodiester backbone of U6 as important for splicing<sup>12,13</sup> (Fig. 1b). Although informative, these studies assayed sulphur substitutions only at the *pro-R<sub>p</sub>* phosphate oxygen (phosphorothioate substitutions referred to as PS(*R<sub>p</sub>*), Fig. 2a) and subsequent studies have only revealed rescue by thiophilic metals for U80-PS(*S<sub>p</sub>*)<sup>14,28</sup>.

To identify metal ligands, we assayed splicing in *Saccharomyces cerevisiae* extracts reconstituted with U6 bearing individual sulphur

<sup>1</sup>Graduate Program in Cell and Molecular Biology, The University of Chicago, Chicago, Illinois 60637, USA. <sup>2</sup>Department of Molecular Genetics and Cell Biology, Cummings Life Sciences Center, 920 East 58th Street, The University of Chicago, Chicago, Illinois 60637, USA. <sup>3</sup>Department of Chemistry, 929 East 57th Street, The University of Chicago, Chicago, Illinois 60637, USA. <sup>4</sup>Department of Biochemistry and Molecular Biology, Gordon Center for Integrative Science, 929 East 57th Street, The University of Chicago, Chicago, Illinois 60637, USA.

\*These authors contributed equally to this work.



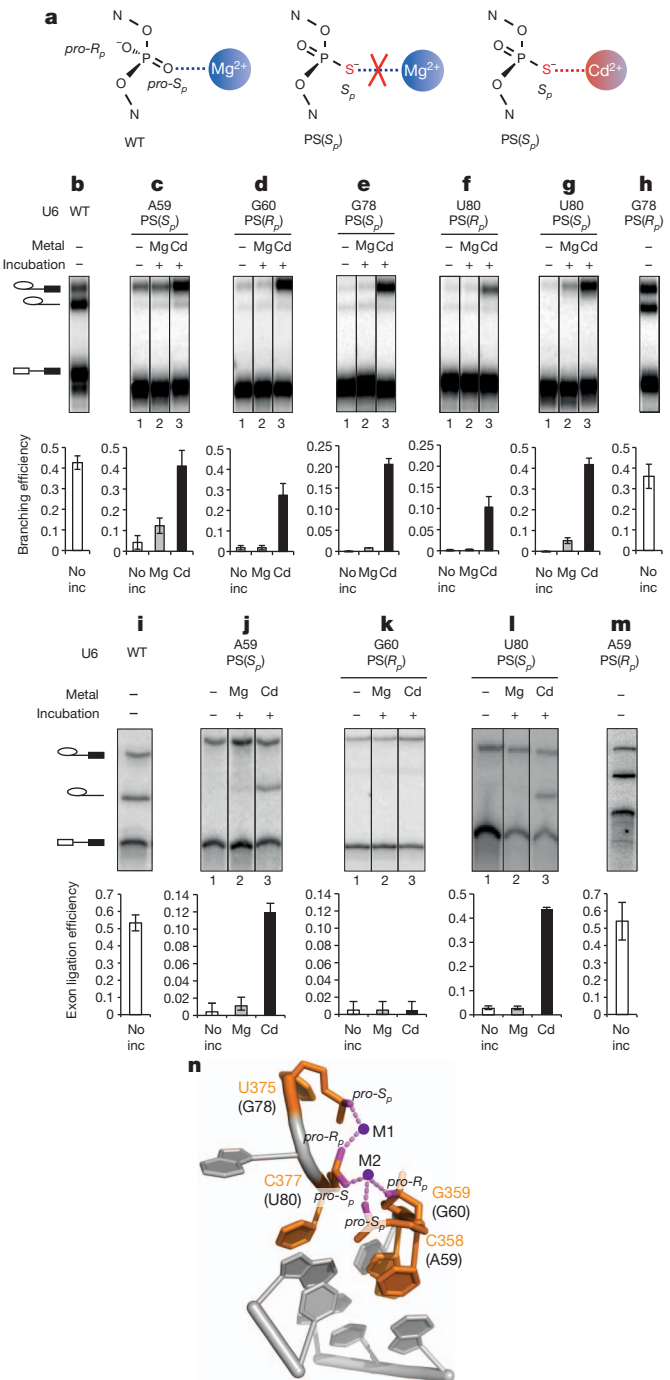
**Figure 1 | Chemistry of pre-mRNA splicing and U2/U6 model showing sites sensitive to sulphur substitutions and rescued by thiophilic metal.** **a**, Reaction scheme (top) and transition state diagrams (bottom) for the two steps of nuclear pre-mRNA splicing. **b**, **c**, Models of the U2/U6 structure from the spliceosome (**b**) and domain V of a group II intron (**c**). Nucleotides where sulphur substitutions were shown previously to interfere with splicing<sup>12–16</sup> are coloured red; underlined U6 nucleotides were tested in this study. An asterisk indicates those positions where only the *pro-S<sub>p</sub>* oxygen was substituted with sulphur; boxed nucleotides were found to provide ligands for metals at the catalytic stage (Fig. 2, ref. 26).

substitutions covering both the *R<sub>p</sub>* and *S<sub>p</sub>* diastereomers at twenty positions, on the basis of previous studies on pre-mRNA and group II splicing<sup>12–16,26</sup> (Fig. 1b, c). To focus on ligand–metal interactions important during catalysis, we restricted our analysis to spliceosomes that had already undergone catalytic activation<sup>36</sup> by affinity-purifying spliceosomes using tagged Prp19p (Extended Data Fig. 1a–c). We then assayed for rescue in the absence of ATP and soluble factors, thus ensuring that spliceosomes had progressed beyond the final ATP-dependent activation step, while at the same time eliminating ATP-dependent proof-reading mechanisms<sup>28,34</sup> to enhance the potential for rescue (Extended Data Fig. 1d, Supplementary Note 1).

Five of the twenty tested substitutions conferred strong branching defects (Fig. 1b). In addition to the sulphur substitutions G60-PS(*R<sub>p</sub>*), U80-PS(*S<sub>p</sub>*), and U80-PS(*R<sub>p</sub>*) (Fig. 2b, h compared with 2d, f, g; first lanes) (refs 12–14), the substitutions G78-PS(*S<sub>p</sub>*) and A59-PS(*S<sub>p</sub>*) also caused branching defects in extract in Mg<sup>2+</sup> (Fig. 2c, e, first lanes; Extended Data Fig. 2a).

In addition to spliceosomes containing U80-PS(*S<sub>p</sub>*) (Fig. 2g; refs 14, 28), affinity-purified spliceosomes containing any of the other four sulphur substitutions that conferred a branching defect could catalyse branching in Cd<sup>2+</sup> much more efficiently than in Mg<sup>2+</sup> (Fig. 2c–f). Thus, all five oxygens sensitive to sulphur substitution bind metals important for branching at the catalytic stage.

To test whether these non-bridging oxygens implicated in branching also function in exon ligation, we assembled each of the twenty sulphur-substituted spliceosomes on a model substrate, chased spliceosomes through branching and assayed for exon ligation after affinity purification to remove ATP. Only spliceosomes containing the five sulphur substitutions implicated in branching failed to catalyse exon ligation in Mg<sup>2+</sup> (Fig. 2i, m compared with 2j–l, first lanes; and data not shown). In addition to U80-PS(*S<sub>p</sub>*) spliceosomes (Fig. 2l; ref. 28),



**Figure 2 | U6 snRNA positions metals important for both steps of splicing.** **a**, Metal rescue strategy, indicating the non-bridging oxygens that were substituted with sulphur and the consequence of a substitution for binding of Mg<sup>2+</sup> or the thiophilic metal Cd<sup>2+</sup>. **b–h**, Cd<sup>2+</sup> rescues the branching defects (b–h) and exon ligation defects (i–m) induced by specific sulphur substitutions. Values represent averages; error bars, s.d. (*n* = 3 for b–h; *n* = 2 for i–m); no inc, no incubation. **n**, Structure of the domain V metal binding core (PDB 3EOH, ref. 26). Residue numbers are shown for the group II intron, with the corresponding U6 residues in parentheses.

A59-PS(*S<sub>p</sub>*) spliceosomes catalysed exon ligation in Cd<sup>2+</sup> (Fig. 2j). Although G60-PS(*R<sub>p</sub>*), G78-PS(*S<sub>p</sub>*) and U80-PS(*R<sub>p</sub>*) failed to catalyse exon ligation even in Cd<sup>2+</sup> (Fig. 2k; data not shown), these sulphur substitutions exhibited more stringent requirements for metal rescue in branching, failing to tolerate the less thiophilic Mn<sup>2+</sup>, whereas U80-PS(*S<sub>p</sub>*) and A59-PS(*S<sub>p</sub>*) spliceosomes did (Extended Data Fig. 2b). Therefore the defects of these sulphur substitutions that could not be

rescued in exon ligation remain consistent with a role for the substituted oxygens as metal ligands during exon ligation (Supplementary Note 2). Thus, at least two of the oxygens that bind a metal important for branching also bind a metal important for exon ligation.

Overall, our analysis shows that five non-bridging oxygens in U6 are bona fide metal ligands in the spliceosome (Fig. 2). These ligands function after catalytic activation, implicating a structural or direct catalytic role for these ligands during the catalytic stage. These ligands correspond directly and stereospecifically to the oxygens that coordinate two divalent metals in domain V of a group II intron, as revealed by X-ray crystallography<sup>26,27</sup> (Fig. 2n). This parallel suggested that the U6 ligands that function at the catalytic stage function analogously to domain V metal ligands by directly positioning the catalytic metals required for splicing.

### U6 positions catalytic metals during branching

Experiments with the *Tetrahymena thermophila* group I intron have established biochemical signatures for identifying ligands that position catalytic metals: sulphur substitution of ligands to a catalytic metal can be rescued more strongly by thiophilic metal or with increased specificity for  $\text{Cd}^{2+}$  when the substrate ligands to that metal are also substituted with sulphur<sup>31</sup>. Consequently, to determine whether the U6 snRNA metal ligands function catalytically during branching, we first identified a substrate sensitive to the identity of catalytic metals required for branching. A pre-mRNA substrate bearing a double sulphur substitution at both the leaving group and the non-bridging *pro-R<sub>p</sub>* oxygen at the 5' splice site (referred to as 3'S-PS(*R<sub>p</sub>*)) was branched efficiently only in the presence of  $\text{Cd}^{2+}$  (Fig. 3a), indicating that in yeast, as in mammals<sup>24</sup> (compare with ref. 37), divalent metals interact with the scissile phosphate (see also Supplementary Notes 3 and 4 and Extended Data Figs 3 and 4). We used the 3'S-PS(*R<sub>p</sub>*) substrate as a reporter for catalytic metal interactions between the spliceosome and the 5' splice site during branching.

These experiments revealed several catalytic interactions between U6 and the 5' splice site. Under sensitized conditions (Fig. 3 legend), the 3'S-PS(*R<sub>p</sub>*) substrate strongly improved  $\text{Cd}^{2+}$ -mediated rescue for G78-PS(*S<sub>p</sub>*) and U80-PS(*R<sub>p</sub>*) spliceosomes (Fig. 3b). In contrast to the

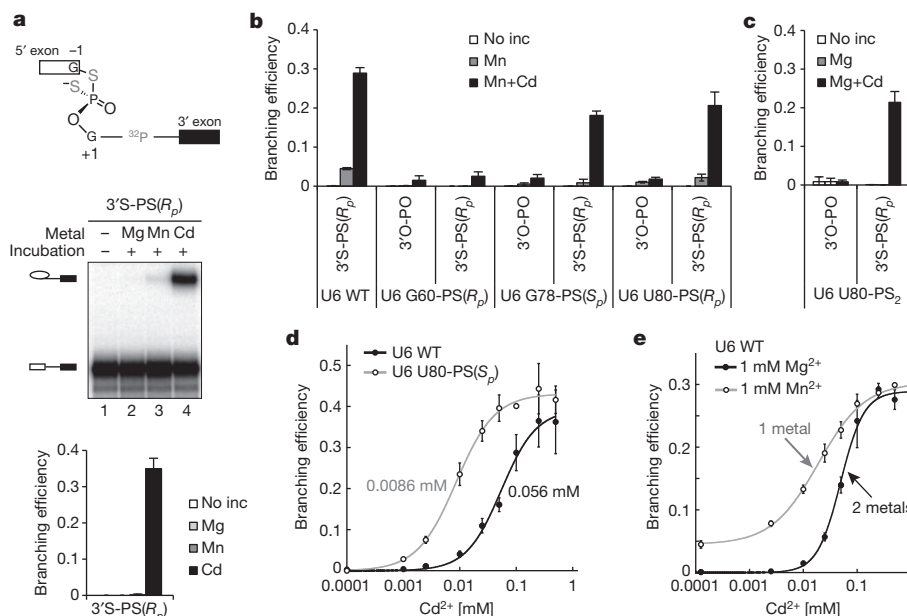
3'S-PS(*R<sub>p</sub>*) substrate, the 3'S-PS(*S<sub>p</sub>*) substrate, having the *pro-S<sub>p</sub>* oxygen, which is not directly involved in catalysis, substituted with sulphur (Extended Data Fig. 3), did not improve  $\text{Cd}^{2+}$ -mediated rescue for G78-PS(*S<sub>p</sub>*) and U80-PS(*R<sub>p</sub>*) (Extended Data Fig. 5a, b). These data indicate that the G78 *pro-S<sub>p</sub>* and U80 *pro-R<sub>p</sub>* oxygens interact with the 5' splice site through catalytic metals during branching.

The 3'S-PS(*R<sub>p</sub>*) substrate also rescued sulphur substitutions that could not be rescued on their own. Substitution of both the *pro-S<sub>p</sub>* and *pro-R<sub>p</sub>* oxygens at U80 with sulphur (U80-PS<sub>2</sub>) impaired branching of the 3'O-PO substrate, but unlike the individual substitutions,  $\text{Cd}^{2+}$  could not rescue branching (Fig. 3c). However, the 3'S-PS(*R<sub>p</sub>*) substrate allowed robust  $\text{Cd}^{2+}$  rescue of branching for U80-PS<sub>2</sub> (Fig. 3c), providing evidence that both of the U80 non-bridging oxygens bind catalytic metals during branching.

The 3'S-PS(*R<sub>p</sub>*) substrate improved rescue of these sulphur substitutions in U6 in a specific manner. The modified substrate did not improve branching for U6 variants that compromised branching due to base mutations (Extended Data Fig. 5c, d, Supplementary Note 5) and did not improve  $\text{Cd}^{2+}$  rescue for G60-PS(*R<sub>p</sub>*) (Fig. 3b). Thus the enhanced rescue of G78-PS(*S<sub>p</sub>*), U80-PS(*R<sub>p</sub>*) and U80-PS<sub>2</sub> spliceosomes conferred specifically by the 3'S-PS(*R<sub>p</sub>*) substrate bears the hallmark of ligands linked by common metals<sup>29–31</sup> and indicates that the corresponding oxygens in U6 position catalytic metals during branching.

An implication of this conclusion is that a sulphur substitution in U6 might also reduce the  $\text{Cd}^{2+}$  concentration required for rescue of a substrate with sulphur substitutions at the 5' splice site, by enhancing  $\text{Cd}^{2+}$  occupancy of a common metal binding site. Indeed, with the 3'S-PS(*R<sub>p</sub>*) substrate, substitution of the *pro-S<sub>p</sub>* oxygen of U80 with sulphur decreased the  $\text{Cd}^{2+}$  titration midpoint for rescue by sixfold compared to wild-type U6 (Fig. 3d, Supplementary Note 6), and at a limiting  $\text{Cd}^{2+}$  concentration (10  $\mu\text{M}$ ), increased the rate of branching sevenfold compared to wild-type U6 (Extended Data Fig. 6a, b), demonstrating that U80-PS(*S<sub>p</sub>*) increased binding of  $\text{Cd}^{2+}$  at the catalytic core.

Our analysis has also revealed that branching requires two distinct catalytic metals (compare with ref. 22). With the 3'S-PS(*R<sub>p</sub>*) substrate,  $\text{Cd}^{2+}$  titrations indicated that rescue required binding of  $\text{Cd}^{2+}$  to two



**Figure 3 | U6 snRNA positions catalytic metals during branching.** **a**, The 3'S-PS(*R<sub>p</sub>*) substrate is sensitive to catalytic metal interactions. **b**, **c**, The 3'S-PS(*R<sub>p</sub>*) substrate improves rescue of G78-PS(*S<sub>p</sub>*), U80-PS(*R<sub>p</sub>*) and U80-PS<sub>2</sub> spliceosomes. **d**, U80-PS(*S<sub>p</sub>*) reduces the  $\text{Cd}^{2+}$  concentration required for branching of the 3'S-PS(*R<sub>p</sub>*) substrate. **e**, The 3'S-PS(*R<sub>p</sub>*) substrate is rescued by

$\text{Cd}^{2+}$  bound at two distinct sites. In **d** and **e**, Hill fits are shown (solid curves); in **e**, titration midpoints are indicated. In **b** and **e**, a lower pH (pH 7.0) was used compared to Fig. 2 (pH 8.5) to sensitize the system for improved rescue. Values represent averages; error bars, s.d. ( $n = 3$ ); no inc, no incubation.



sites, when  $\text{Cd}^{2+}$  was the only thiophilic metal present (Fig. 3e). In contrast, when  $\text{Mn}^{2+}$  was also present, rescue still required binding of  $\text{Cd}^{2+}$  but to only one site (Fig. 3e; Extended Data Fig. 7a–c), demonstrating that one metal site could bind  $\text{Mn}^{2+}$  whereas the other site could not. Given that  $\text{Cd}^{2+}$  is expected to bind a site containing several sulphur ligands, whereas  $\text{Mn}^{2+}$  is not<sup>38</sup>, these and additional data indicate that one metal interacts with both sulphur atoms of the 3'S-PS( $R_p$ ) substrate (referred to as M1) (Extended Data Fig. 7e, f), whereas the second metal only interacts with the non-bridging  $R_p$  sulphur (M2). Further evidence indicates that the U80 *pro-R<sub>p</sub>* and G78 *pro-S<sub>p</sub>* oxygens also interact with the M1 site, whereas the U80 *pro-S<sub>p</sub>* oxygen interacts with the second metal site (M2) (Extended Data Figs 7 and 8, Supplementary Note 7). Thus, at least three of the five identified U6 ligands coordinate two distinct catalytic metals that interact with the scissile phosphate of the 5' splice site during branching (Supplementary Note 8, Extended Data Figs 5e–g and 6e).

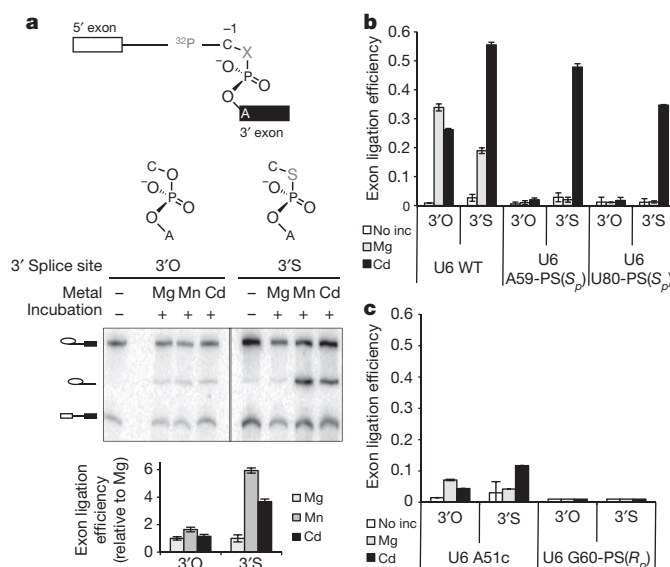
### U6 positions a catalytic metal during exon ligation

To determine whether the U6 metal ligands function during exon ligation by directly binding catalytic metals, we tested for biochemical signatures<sup>31</sup> as above, that would link the U6 ligands to the catalytic metal that interacts with the 3' splice site during exon ligation.

First, we introduced a sulphur substitution at the 3' splice site leaving group to sensitize splicing of a model substrate to binding of the catalytic metal that interacts with this group. Additionally, we introduced a mutation of the 3' splice site consensus sequence (UAG to UAC) to stall spliceosomes before exon ligation<sup>34</sup> regardless of whether the 3'-oxygen leaving group at the 3' splice site was substituted with sulphur or oxygen (3'S or 3'O). Exon ligation of this substrate can proceed only in the absence of ATP due to proofreading<sup>34</sup>. Although exon ligation of affinity-purified UAC-3'O spliceosomes proceeded in the absence of ATP, the 3'S substitution compromised exon ligation when only  $\text{Mg}^{2+}$  was present (Fig. 4a). When  $\text{Mn}^{2+}$  or  $\text{Cd}^{2+}$  was present, spliceosomes assembled on the UAC-3'S substrate catalysed exon ligation fourfold to sixfold more efficiently (Fig. 4a), indicating that a divalent metal stabilizes the leaving group in yeast, as in mammals<sup>25</sup> (see also Supplementary Note 9 and Extended Data Fig. 9a–f). Because the UAC-3'S substrate specifically required thiophilic, catalytic metals, we used it as a reporter for catalytic metal interactions between the spliceosome and the 3' splice site during exon ligation (Supplementary Note 10).

After assembling spliceosomes containing U6 sulphur substitutions on this sulphur-substituted substrate, we probed for exon ligation at the catalytic stage (that is, in the absence of ATP and soluble factors). As expected from splicing assays of the wild-type UAG-3'O substrate (Fig. 2j, l), exon ligation of the mutated UAC-3'O substrate in  $\text{Mg}^{2+}$  occurred with exceedingly low efficiency with spliceosomes containing U6 A59-PS( $S_p$ ) or U80-PS( $S_p$ ) (Fig. 4b). However, in contrast to exon ligation assays of the wild-type UAG-3'O substrate (Fig. 2j, l), exon ligation of the mutated UAC-3'O substrate by these sulphur-substituted spliceosomes was not rescued by  $\text{Cd}^{2+}$  (Fig. 4b), indicating that  $\text{Cd}^{2+}$  could no longer rescue the U6 sulphur substitutions in the context of the substrate base mutation. Nevertheless, when the substrate also contained sulphur in the leaving group position (UAC-3'S),  $\text{Cd}^{2+}$ , but not  $\text{Mg}^{2+}$ , strongly stimulated exon ligation by the sulphur-substituted spliceosomes (Fig. 4b, Extended Data Fig. 9g). This stronger rescue by  $\text{Cd}^{2+}$  of a sulphur-substituted ligand in the presence of a second sulphur-substituted ligand constitutes a signature for a functional and physical link between two ligands that bind the same metal<sup>29–31</sup>. These data therefore establish that the U6 A59 *pro-S<sub>p</sub>* and U80 *pro-S<sub>p</sub>* oxygens interact with the catalytic metal that stabilizes the leaving group during exon ligation.

The signature was specific to the U6 A59-PS( $S_p$ ) or U80-PS( $S_p$ ) spliceosomes and was not observed for other spliceosomes that were compromised for exon ligation. Unlike the U6 sulphur-substituted spliceosomes assembled on the UAC-3'S substrate, the exon ligation



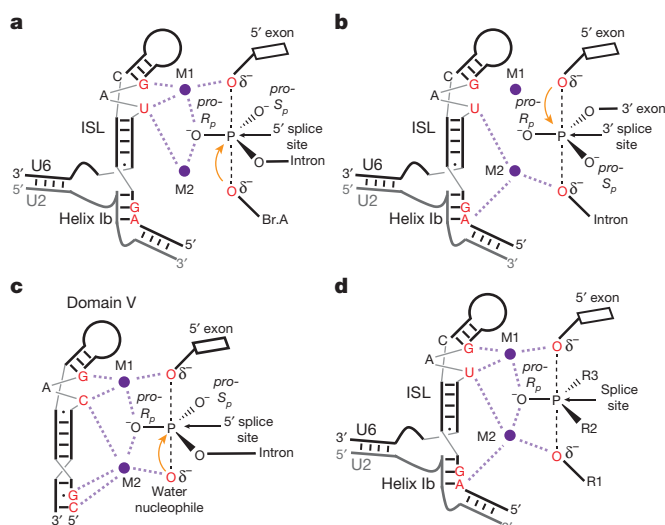
**Figure 4 | U6 snRNA positions a catalytic metal during exon ligation.** **a**, The UAC-3'S substrate is sensitive to catalytic metal interactions. **b**, **c**, The UAC-3'S substrate specifically improves rescue of U6 A59-PS( $S_p$ ) and U80-PS( $S_p$ ) spliceosomes. Values represent averages; error bars, s.d. ( $n = 3$ ); no inc, no incubation.

defect<sup>39</sup> of U6 A51c spliceosomes was not substantially suppressed in  $\text{Cd}^{2+}$  by the UAC-3'S substrate (Fig. 4c). Furthermore, the exon ligation defect of G60-PS( $R_p$ ) was not suppressed by the UAC-3'S substrate in  $\text{Cd}^{2+}$  (Fig. 4c; Extended Data Fig. 9h; Supplementary Note 11), indicating that the substrate sulphur substitution alone does not enable exon ligation of compromised spliceosomes. These results and a metal specificity switch induced by U80-PS( $S_p$ ) (Extended Data Fig. 8 and Supplementary Note 12) validate the evidence that the U6 A59 *pro-S<sub>p</sub>* and U80 *pro-S<sub>p</sub>* oxygens interact with a catalytic metal during exon ligation.

Through a comprehensive analysis, encompassing every substrate allowed by sulphur chemistry (Extended Data Figs 5g and 9h), we have found that four of the five metal ligands in U6 position catalytic metals during the catalytic stages of splicing. Even though metal rescue strategies were insufficient to establish direct evidence linking the fifth ligand (G60 *pro-S<sub>p</sub>*) to a splice site (Supplementary Note 2), the configuration of the RNA catalytic core implied by our data are consistent with the fifth ligand also binding a catalytic metal (see below). Importantly, at least two of the U6 metal ligands function in both steps of splicing and at least one of these, U80 *pro-S<sub>p</sub>*, positions a catalytic metal during both branching and exon ligation (M2, Fig. 5a, b). These and other findings (Supplementary Note 13) support a model<sup>22</sup> in which both splicing reactions are catalysed by a common, two-metal catalytic core, rather than two independent active sites, and thereby implicate a rearrangement of the substrate to sequentially accommodate mutually exclusive interactions between U6 and the 5' and 3' splice sites. Further underscoring the importance of disabling fidelity mechanisms in our approach, these results establish that the spliceosome proofreads catalytic interactions (compare with ref. 28).

### Discussion

Our results indicate that the metals that mediate splicing catalysis through interactions with the scissile phosphates are bound by the spliceosome through ligands in the U6 snRNA, thus demonstrating that RNA directly mediates catalysis in the spliceosome (Fig. 5d). Notably, all of the five U6 metal ligands that function in pre-mRNA splicing (Fig. 2) correspond directly to catalytic metal ligands observed in domain V in structures of a group II intron (Figs 1b, c; 2n; 5c; refs 26, 27). Moreover, our findings imply an equivalent orientation of the substrate



**Figure 5 | Model for catalytic metal interactions during pre-mRNA splicing and comparison to the domain V catalytic core of group II introns.**

**a, b,** Catalytic metal interactions during branching (**a**) and exon ligation (**b**). **c,** Model of domain V during hydrolysis (PDB 4FAQ, ref. 27). **d,** Two-metal model for the RNA catalytic core of the spliceosome. For branching, R1 represents the 2' hydroxyl of the branch adenosine; R2, the intron; and R3, the *pro-S<sub>p</sub>* oxygen. For exon ligation R1 represents the 3' oxygen leaving group, R2 the *pro-S<sub>p</sub>* oxygen; and R3 the 3' exon. Throughout the reactive oxygens are coloured red, the pre-mRNA scissile phosphate is depicted in a transition state, and interactions between specific ligands and the reactive oxygens mediated by M1 and M2 are shown as light magenta dashed lines.

relative to the metal binding core (Fig. 5a–c). Taken together, our data support a model for a single, two-metal active site (Fig. 5d; Supplementary Note 13; refs. 22, 40) and validate (compare with ref. 25) the long-standing hypothesis that the spliceosome shares a similar RNA-based catalytic core and mechanism with group II introns<sup>5,6,13,18,22,27,41</sup> (Supplementary Note 14). This RNA-based mechanism is sufficient to effect metal catalysis of pre-mRNA splicing, without the need for direct protein involvement.

Given this evidence, it is noteworthy that the RNaseH-like domain of Prp8, which interacts with all reactive sites of the substrate, can bind a metal *in crystallo*<sup>42</sup>. Further, mutations that compromise metal binding *in crystallo* impair exon ligation, leading to the possibility that Prp8 may play a catalytic role during exon ligation<sup>42</sup>. Nevertheless, we have not found any evidence for a direct metal-mediated catalytic interaction between the Prp8 metal binding site and the 3' splice site (Supplementary Note 15; Extended Data Fig. 10). Thus, Prp8 may primarily promote formation of the RNA catalytic core without being a part of it; indeed, a recent structure comprising most of Prp8 has led to the suggestion that Prp8 forms a scaffold for the spliceosome's catalytic core, analogously to protein co-factors of group II introns<sup>17</sup>. A recent model of the catalytic core of the human spliceosome based on RNA structure probing and homology modelling provides further support for this interpretation<sup>40,43</sup>.

Overall, our data indicate that the spliceosome, like the ribosome<sup>44,45</sup>, uses RNA to effect catalysis in the context of a complex ribonucleoprotein assembly. Moreover, the common catalytic mechanism used by the spliceosome and group II introns is consistent with a common evolutionary origin between the spliceosome and these ancient RNA retroelements<sup>46,47</sup>. Our findings thus support the idea that modern ribonucleoprotein enzymes evolved from a primordial 'RNA world' (ref. 48), in which catalysis was performed exclusively by RNA.

## METHODS SUMMARY

U6 was depleted from *Saccharomyces cerevisiae* splicing extracts and splicing activity reconstituted with synthetic U6 snRNA, essentially as described<sup>28</sup>. Spliceosomes

were assembled on modified model pre-mRNA substrates: *UBC4*, for experiments probing branching, or *ACT1*, for experiments probing exon ligation, both synthesized by splint-mediated ligation<sup>49</sup>. Oligonucleotides containing specific 5' or 3' splice site modifications were synthesized in house, as described previously<sup>50</sup>. Assembled spliceosomes were isolated by affinity purification via Prp19p (ref. 36), washed to remove ATP, and chased as described<sup>28</sup>, at room temperature, in the absence of ATP, at pH 7.0, 8.0 or 8.5. All experiments were repeated with at least two independent extract preparations. Data were quantified using ImageQuant TL (Amersham Biosciences).

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 May; accepted 3 October 2013.

Published online 6 November 2013.

- Nilsen, T. W. & Graveley, B. R. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463**, 457–463 (2010).
- Wahl, M. C., Will, C. L. & Lührmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701–718 (2009).
- Sharp, P. A. On the origin of RNA splicing and introns. *Cell* **42**, 397–400 (1985).
- Cech, T. R. The generality of self-splicing RNA: relationship to nuclear mRNA splicing. *Cell* **44**, 207–210 (1986).
- Madhani, H. D. & Guthrie, C. A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell* **71**, 803–817 (1992).
- Shukla, G. C. & Padgett, R. A. A catalytically active group II intron domain 5 can function in the U2-dependent spliceosome. *Mol. Cell* **9**, 1145–1150 (2002).
- Hilliker, A. K. & Staley, J. P. Multiple functions for the invariant AGC triad of U6 snRNA. *RNA* **10**, 921–928 (2004).
- Mefford, M. A. & Staley, J. P. Evidence that U2/U6 helix I promotes both catalytic steps of pre-mRNA splicing and rearranges in between these steps. *RNA* **15**, 1386–1397 (2009).
- Burke, J. E., Sashital, D. G., Zuo, X., Wang, Y.-X. & Butcher, S. E. Structure of the yeast U2/U6 snRNA complex. *RNA* **18**, 673–683 (2012).
- Sun, J. S. & Manley, J. L. A novel U2–U6 snRNA structure is necessary for mammalian mRNA splicing. *Genes Dev.* **9**, 843–854 (1995).
- Chanfreau, G. & Jacquier, A. Catalytic site components common to both splicing steps of a group II intron. *Science* **266**, 1383–1387 (1994).
- Fabrizio, P. & Abelson, J. Thiophosphates in yeast U6 snRNA specifically affect pre-mRNA splicing *in vitro*. *Nucleic Acids Res.* **20**, 3659–3664 (1992).
- Yu, Y. T., Maroney, P. A., Darzynkiewicz, E. & Nilsen, T. W. U6 snRNA function in nuclear pre-mRNA splicing: a phosphorothioate interference analysis of the U6 phosphate backbone. *RNA* **1**, 46–54 (1995).
- Yean, S. L., Wuenschell, G., Termini, J. & Lin, R.-J. Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. *Nature* **408**, 881–884 (2000).
- Gordon, P. M. & Piccirilli, J. A. Metal ion coordination by the AGC triad in domain 5 contributes to group II intron catalysis. *Nature Struct. Biol.* **8**, 893–898 (2001).
- Boudvillain, M. & Pyle, A. M. Defining functional groups, core structural features and inter-domain tertiary contacts essential for group II intron self-splicing: a NAIM analysis. *EMBO J.* **17**, 7091–7104 (1998).
- Galej, W. P., Oubridge, C., Newman, A. J. & Nagai, K. Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature* **493**, 638–643 (2013).
- Valadkhan, S. & Manley, J. L. Splicing-related catalysis by protein-free snRNAs. *Nature* **413**, 701–707 (2001).
- Jaladat, Y., Zhang, B., Mohammadi, A. & Valadkhan, S. Splicing of an intervening sequence by protein-free human snRNAs. *RNA Biol.* **8**, 372–377 (2011).
- Smith, D. J. & Konarska, M. M. A critical assessment of the utility of protein-free splicing systems. *RNA* **15**, 1–3 (2009).
- Valadkhan, S. & Manley, J. L. The use of simple model systems to study spliceosomal catalysis. *RNA* **15**, 4–7 (2009).
- Steitz, T. A. & Steitz, J. A. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl Acad. Sci. USA* **90**, 6498–6502 (1993).
- Sontheimer, E. J., Gordon, P. M. & Piccirilli, J. A. Metal ion catalysis during group II intron self-splicing: parallels with the spliceosome. *Genes Dev.* **13**, 1729–1741 (1999).
- Sontheimer, E. J., Sun, S. & Piccirilli, J. A. Metal ion catalysis during splicing of premessenger RNA. *Nature* **388**, 801–805 (1997).
- Gordon, P. M., Sontheimer, E. J. & Piccirilli, J. A. Metal ion catalysis during the exon-ligation step of nuclear pre-mRNA splicing: extending the parallels between the spliceosome and group II introns. *RNA* **6**, 199–205 (2000).
- Toor, N., Keating, K. S., Taylor, S. D. & Pyle, A. M. Crystal structure of a self-spliced group II intron. *Science* **320**, 77–82 (2008).
- Marcia, M. & Pyle, A. M. Visualizing group II intron catalysis through the stages of splicing. *Cell* **151**, 497–507 (2012).
- Koodathingal, P., Novak, T., Piccirilli, J. A. & Staley, J. P. The DEAH box ATPases Prp16 and Prp43 cooperate to proofread 5' splice site cleavage during pre-mRNA splicing. *Mol. Cell* **39**, 385–395 (2010).
- Shan, S., Yoshida, A., Sun, S., Piccirilli, J. A. & Herschlag, D. Three metal ions at the active site of the *Tetrahymena* group I ribozyme. *Proc. Natl Acad. Sci. USA* **96**, 12299–12304 (1999).

30. Forconi, M., Lee, J., Lee, J. K., Piccirilli, J. A. & Herschlag, D. Functional identification of ligands for a catalytic metal ion in group I introns. *Biochemistry* **47**, 6883–6894 (2008).
31. Frederiksen, J. K. & Piccirilli, J. A. Identification of catalytic metal ion ligands in ribozymes. *Methods* **49**, 148–166 (2009).
32. Guo, F., Gooding, A. R. & Cech, T. R. Structure of the *Tetrahymena* ribozyme: base triple sandwich and metal ion at the active site. *Mol. Cell* **16**, 351–362 (2004).
33. Burgess, S. M. & Guthrie, C. A mechanism to enhance mRNA splicing fidelity: the RNA-dependent ATPase Prp16 governs usage of a discard pathway for aberrant lariat intermediates. *Cell* **73**, 1377–1391 (1993).
34. Mayas, R. M., Maita, H. & Staley, J. P. Exon ligation is proofread by the DExD/H-box ATPase Prp22p. *Nature Struct. Mol. Biol.* **13**, 482–490 (2006).
35. Semlow, D. R. & Staley, J. P. Staying on message: ensuring fidelity in pre-mRNA splicing. *Trends Biochem. Sci.* **37**, 263–273 (2012).
36. Chan, S.-P., Kao, D.-I., Tsai, W.-Y. & Cheng, S.-C. The Prp19p-associated complex in spliceosome activation. *Science* **302**, 279–282 (2003).
37. Moore, M. J. & Sharp, P. A. Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature* **365**, 364–368 (1993).
38. Pecoraro, V. L., Hermes, J. D. & Cleland, W. W. Stability constants of  $Mg^{2+}$  and  $Cd^{2+}$  complexes of adenine nucleotides and thionucleotides and rate constants for formation and dissociation of MgATP and MgADP. *Biochemistry* **23**, 5262–5271 (1984).
39. Fabrizio, P. & Abelson, J. J. Two domains of yeast U6 small nuclear RNA required for both steps of nuclear precursor messenger RNA splicing. *Science* **250**, 404–409 (1990).
40. Anokhina, M. et al. RNA structure analysis of human spliceosomes reveals a compact 3D arrangement of snRNAs at the catalytic core. *EMBO J.* **32**, 2804–2818 (2013).
41. Cech, T. R. The chemistry of self-splicing RNA and RNA enzymes. *Science* **236**, 1532–1539 (1987).
42. Schellenberg, M. J. et al. A conformational switch in PRP8 mediates metal ion coordination that promotes pre-mRNA exon ligation. *Nature Struct. Mol. Biol.* **20**, 728–734 (2013).
43. Bonnal, S. & Valcárcel, J. RNAtomy of the spliceosome's heart. *EMBO J.* **32**, 2785–2787 (2013).
44. Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930 (2000).
45. Beringer, M. & Rodnina, M. V. The ribosomal peptidyl transferase. *Mol. Cell* **26**, 311–321 (2007).
46. Sharp, P. A. Five easy pieces. *Science* **254**, 663 (1991).
47. Martin, W. & Koonin, E. V. Introns and the origin of nucleus–cytosol compartmentalization. *Nature* **440**, 41–45 (2006).
48. Joyce, G. F. The antiquity of RNA-based evolution. *Nature* **418**, 214–221 (2002).
49. Moore, M. J. & Sharp, P. Site-specific modification of pre-mRNA: the 2'-hydroxyl groups at the splice sites. *Science* **256**, 992–997 (1992).
50. Yoshida, A., Sun, S. & Piccirilli, J. A. A new metal ion interaction in the *Tetrahymena* ribozyme reaction revealed by double sulfur substitution. *Nature Struct. Biol.* **6**, 318–321 (1999).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank C. Guthrie for plasmids; S.-C. Cheng for anti-Cwc25p serum; D. Semlow for strains; J. Olvera for reagents and experimental assistance; R.-J. Lin for sharing unpublished data; members of the Staley and Piccirilli laboratories for discussions; and D. Herschlag, A. Macmillan and T. Nilsen for comments on the manuscript. N.T. was supported by an NSF Graduate Research Fellowship and by a CBI Training Grant (5T32GM008720). This work was funded by a grant from the Chicago Biomedical Consortium, with support from The Searle Funds at the Chicago Community Trust, to J.P.S., A. S. Mankin and E. J. Sontheimer, and by a grant from the National Institutes of Health (R01GM088656) to J.P.S. and J.A.P.

**Author Contributions** S.M.F., N.T., T.N., J.P.S. and J.A.P. designed the study; T.N. and P.K. performed initial screening of U6 sulphur substitutions; S.M.F. performed all experiments related to branching; N.T. performed all experiments related to exon ligation; S.M.F. and N.T. together performed Prp8p experiments; J.L., N.-S.L. and Q.D. synthesized RNA oligonucleotides; S.M.F., N.T., J.P.S. and J.A.P. analysed the data and wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.A.P. ([jpiccirilli@uchicago.edu](mailto:jpiccirilli@uchicago.edu)) or J.P.S. ([jstaley@uchicago.edu](mailto:jstaley@uchicago.edu)).



## METHODS

**Strains.** Experiments probing branching were performed using *S. cerevisiae* strain yJPS1405, which was derived from BY4741 (*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*, Open Biosystems), by integrating a biotinylation signal C-terminal to the *PRP19* locus using *KanMX6* as a selective marker; the integration fragment, including the biotinylation signal, an *ADH1* terminator, and the *KanMX* sequences were amplified from pFA6-HTB-*KanMX* (ref. 51).

Experiments probing exon ligation were performed using the previously described *S. cerevisiae* strain yJPS860 containing a C-terminal TAP tag on Prp19 (ref. 34).

Experiments involving Prp8 were performed using *S. cerevisiae* strains yJPS1471 (*PRP8* wild-type) and yJPS1472 (*PRP8* D1853C), both derived from yJPS860 (ref. 34). Briefly, the entire coding sequence of *PRP8* was replaced in yJPS860 with *LEU2*, after transformation with *PRP8* on a *URA3* plasmid (bJPS1874), yielding yJPS1470. Then yJPS1471 and 1472 were derived from yJPS1470 by plasmid shuffle on 5-FOA after transformation with bJPS2263 (*PRP8* wild-type on a *HIS3* plasmid, same as pJU186, ref. 52) or bJPS2589 (*PRP8* D1853C). bJPS2589 was derived from pJU186 (ref. 52) by QuikChange mutagenesis and confirmed by sequencing.

**Pre-mRNA splicing substrates.** Branching experiments were performed using *UBC4* pre-mRNA<sup>53</sup> truncated at the 5' and 3' ends down to 20 nucleotide exons. Exon ligation experiments were performed using *ACT1* pre-mRNA. Pre-mRNA substrates were prepared by splint-mediated ligation<sup>49</sup> using T4 RNA ligase 2 (NEB) or T4 DNA ligase (in house).

Oligonucleotides UBC4-5'E, 5'E2, UBC4-M, and UBC4-M1 were purchased from Dharmacon. Oligonucleotide UBC4-M2, bearing various modifications, was synthesized in house<sup>50</sup> and purified by HPLC on a DNAPac P-100 column using the Waters 2795 system<sup>54</sup>. Oligonucleotides ACT1-3'O-pc, ACT1-3'S-pc, ACT1-UAc-3'O, and ACT1-UAc-3'S were synthesized in house. The correct identity (position and chirality) of specific sulphur modifications was verified by mass spectrometry, silver or iodine cleavage, and analytical T1 digestion<sup>55</sup> (data not shown). Oligos M, M1, and M2 were phosphorylated with unlabelled ATP before ligation. The UBC4-TX transcript was synthesized by *in vitro* transcription using a PCR-derived template that started transcription at position +37 of the intron and ended at the last nucleotide of the exon; the template was generated using primers T7 UBC4 37-135 F and HindIII UBC4 R. The transcript was gel purified, treated with calf inorganic phosphatase (NEB), and 5' phosphorylated with [ $\gamma$ -<sup>32</sup>P]ATP (Perkin Elmer, 6,000 C mmol<sup>-1</sup>) before ligation.

The ACT1-5'-end plasmid was constructed by stepwise PCR from plasmid pJPS149 (ref. 34) to generate a DNA template containing an EcoRI site, the T7 promoter, ACT1 nucleotides 1–373, an HDV ribozyme sequence<sup>56</sup> and a HindIII site. The template was then cloned into vector pUC19. The ACT1-1-373 transcript was synthesized by *in vitro* transcription from the ACT1-5'-piece plasmid linearized with HindIII. In cases where HDV cleavage was inefficient during transcription, the RNA was resuspended in 10 mM Tris (pH 7.5) and 20 mM MgCl<sub>2</sub>. Ribozyme cleavage was induced by 2–4 cycles of 90 °C for 1 min, room temperature for 15 min and 37 °C for 15 min. The buffer conditions were then adjusted for T4 PNK (NEB) treatment of the transcript to remove the 2'-3'-cyclic phosphate left by the ribozyme. The ACT1-392–590 transcript was synthesized by *in vitro* transcription using a PCR-derived template generated using plasmid bJPS149 (ref. 34). As the subsequent ligation requires a 5'-monophosphate group, a fourfold excess of GMP over GTP was included in the transcription reaction.

For a typical *UBC4* ligation reaction, 20 pMol of UBC4 5'E, 10 pMol of UBC4 M, M1, or M2 and 20 pMol of UBC4 TX were hybridized to 10 pMol of UBC4 splint in buffer TEN (10 mM Tris-HCl, pH 7.5; 1 mM EDTA; 66 mM NaCl) on a thermal cycler by heating to 90 °C for 2 min followed by reduction of the temperature by 1 °C for 1 min in 72 sequential steps, to a final temperature of 18 °C. T4 RNA ligase 2 (10–20 U) was then added and reactions were incubated at 37 °C for 6 h. Ligated, full-length *UBC4* pre-mRNA was purified on an 8% denaturing polyacrylamide gel and recovered by passive elution in TE buffer at 4 °C overnight (10 mM Tris-HCl pH 7.5; 1 mM EDTA).

For a typical *ACT1* ligation reaction, 500 pMol of ACT1-1-373, 50 pMol of oligonucleotide ACT1-3'O and 500 pMol of ACT1-392–590 were hybridized to 50 pMol of ACT1 splint in buffer TEN<sub>50</sub> (10 mM Tris-HCl, pH 7.5; 1 mM EDTA; 50 mM NaCl) on a thermal cycler by heating to 90 °C for 2 min followed by reduction of the temperature by 1 °C for 1 min to 24 °C, then cooling to 4 °C for 5 min. T4 DNA ligase (~100 pMol, synthesized in house) was then added and reactions were incubated at 37 °C for 4 h. The ligation reactions were DNase treated (RNase-free DNase, Promega) for 15 min to remove splint, phenol-chloroform extracted and ethanol precipitated before purification on 6% denaturing polyacrylamide gel. Bands containing full-length *ACT1* pre-mRNA were excised and recovered by passive elution in TEN<sub>250</sub> buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 250 mM NaCl) overnight at 4 °C.

**Splicing extracts and U6 depletion and reconstitution.** Splicing extracts were prepared using the liquid nitrogen method, as described<sup>52</sup>. U6 depletion and reconstitution was performed essentially as described<sup>28</sup>, with the following modifications. The U6 d1 oligonucleotide was titrated for each extract to optimize depletion and reconstitution of U6; typically, 0.8 μM U6 d1 was optimal. After depletion of U6, to promote degradation of the U6 d1 oligonucleotide before addition of the synthetic U6, DNase I (Ambion, 0.05 U μl<sup>-1</sup> final) was added together with the glucose used to deplete ATP. Additionally, to ensure complete inactivation of the U6 d1 oligonucleotide, an oligonucleotide antisense to U6 d1 (U6 αd1, 0.3 μM final)<sup>57</sup> was added immediately before reconstitution and incubated on ice for 5 min. For reconstitution U6 was added back to a final concentration of 0.2–0.3 μM. Modified U6 was constructed by splint-mediated ligation essentially as described<sup>28</sup>. Wild-type U6 was synthesized by *in vitro* transcription according to standard procedures using pJPS488 linearized with DraI as template<sup>58</sup>. There was no difference in splicing efficiency between spliceosomes reconstituted with wild type U6 made by ligation or by *in vitro* transcription (data not shown and Extended Data Fig. 6b).

***In vitro* splicing.** *In vitro* splicing was performed essentially as described<sup>28</sup> using <sup>32</sup>P-labelled substrates (0.2–0.4 nM). In Fig. 2c, g, extracts were pre-incubated with 2 mM EDTA at 4 °C for 30 min before assembling the reactions; MgCl<sub>2</sub> was adjusted to 3.3 mM during the splicing reaction (compared to standard 2.5 mM) to compensate for the final EDTA concentration (0.8 mM). Reactions were usually incubated in extract for 25 min before affinity purification of assembled complexes, except for experiments in Fig. 3a and Extended Data Fig. 3, where initial incubation was performed for 10 min. Assembled and stalled complexes were affinity-purified, via tagged Prp19p, by incubating the splicing reactions for 1–3 h at 4 °C with a 10%–20% reaction volume of streptavidin-agarose (Thermo Scientific) or IgG-Sepharose (GE) slurry pre-washed twice with 25–50 volumes of IPP<sub>150</sub> (10 mM Tris-HCl pH 8.0; 150 mM NaCl; 0.01% NP-40 substitute (Fluka)). Following immunoprecipitation, beads were washed at 4 °C twice with 50 volumes of buffer PK (3% PEG<sub>8000</sub>, 60 mM potassium phosphate, pH 7.0), and spliceosomes were assayed for splicing in the absence of ATP (except where noted), in buffer PK (pH 7.0, 8.0, or 8.5, as indicated) with various amounts of the indicated metals and EDTA (where noted) at room temperature (23–24 °C) with constant rotation, for 60 to 90 min. Metal solutions were prepared from corresponding powdered salts of at least 99.99% purity (Sigma).

Affinity-purified spliceosomes were incubated as follows. In Figs 2c and 3a – buffer PK (pH 8.5, with 0.5 mM EDTA) with 1 mM MgCl<sub>2</sub> and 1 mM of the indicated metals; in Fig. 2d–g, buffer PK (pH 8.5) with 0.5 mM MgCl<sub>2</sub> and 0.5 mM of the indicated metals; in Figs 2j–l and 4b, c, buffer PK (pH 7.0) with 2 mM MgCl<sub>2</sub> and 0.1 mM CdCl<sub>2</sub>; in Fig. 2k, l spliceosomes containing A59-PS(*S<sub>p</sub>*) and U80-PS(*S<sub>p</sub>*) were chased through branching in the presence of 0.1 mM CdCl<sub>2</sub> before affinity purification. In Fig. 3b, e, buffer PK (pH 7.0) with 1 mM MnCl<sub>2</sub> alone or in combination with 1 mM CdCl<sub>2</sub>; in Fig. 3c, buffer PK (pH 8.5) with 0.5 mM MgCl<sub>2</sub> and 0.5 mM CdCl<sub>2</sub>; in Fig. 3d, buffer PK (pH 8.5) with 0.5 mM MgCl<sub>2</sub> and various amounts of CdCl<sub>2</sub>, as indicated. In Fig. 4a, buffer PK (pH 7.0) with 1 mM MgCl<sub>2</sub> and 0.1 mM CdCl<sub>2</sub> or 0.1 mM MnCl<sub>2</sub>.

Following *in vitro* splicing, products were separated on 15% denaturing polyacrylamide gels for *UBC4* substrates or 6% denaturing polyacrylamide gels for *ACT1* substrates. All experiments were repeated with at least two independent extract preparations.

**Cwc25p depletion and reconstitution.** For depletion of Cwc25p, splicing extracts were prepared fresh, dialysed against buffer D (ref. 52), and then incubated at 4 °C for 1.5 h with 50% volume equivalent of protein A-sepharose (Sigma) (1:1 slurry in IPP<sub>150</sub>) conjugated to anti-Cwc25 serum (gift from S. -C. Cheng, ref. 59) and pre-equilibrated in buffer D by rotation at 4 °C for 15 min. Beads were removed by centrifugation at 800g for 4 min and the supernatant was used as depleted extract.

rCwc25p extended with a C-terminal six-histidine tag was expressed in *E. coli* BL21(DE3)pLysS transformed with pET15b (ref. 59; a gift from S. -C. Cheng). Induction in *E. coli* and purification by Ni<sup>2+</sup>-NTA affinity chromatography were performed essentially as described<sup>60</sup> but binding and washing was performed manually. Following elution from the Ni<sup>2+</sup>-NTA resin, the protein was further purified by glycerol gradient centrifugation to more than 90% purity (as estimated by Coomassie blue staining).

To prepare the heat-soluble extract fraction (HP), yeast splicing extracts were incubated at 90 °C for 5 min and insoluble material was removed by centrifugation at 16000g for 5 min. The supernatant was used as HP.

For complementation of affinity purified spliceosomes from extracts immunodepleted of Cwc25p, spliceosomes were washed twice with 50 volumes of buffer DK (20 mM HEPES-KOH, pH 7.9; 60 mM potassium phosphate, pH 7.0; 50 mM NaCl; 0.2 mM EDTA) and splicing was assayed in buffer DK, in the presence of rCwc25-6His (0.7–1.4 μM) and a heat soluble extract fraction (HP; typically 1–2 μl for a 50 μl splicing reaction).

To determine the rates of branching in Extended Data Fig. 4d, reactions were pre-incubated with rCwc25p and HP in the absence of metal at room temperature for 5 min to allow binding and temperature equilibration. Splicing was initiated by addition of metals; aliquots were removed at various times, immediately quenched in STOP buffer (50 mM NaOAc, pH 5.2; 1 mM EDTA; 0.1% SDS; 0.1 mg ml<sup>-1</sup> glycogen) and placed on ice before phenol extraction.

**Oligonucleotides.** The following oligonucleotides were used for synthesis of UBC4 substrates. UBC4 5'E: 5'-GAACUAAGUGAUCUAGAAAGG-3'; UBC4 5'Ev2: 5'-GAACUAAGUGAUCUA-3'; UBC4 M: 5'-UAUGUCUAAAGUU AU-3'; UBC4 M1: 5'-GAAAG(3'S)GUAUGUCUAAAGUU AU-3'; UBC4 M2: 5'-GAAAG(3'S-PS)GUAUGUCUAAAGUU AU-3'; UBC4 Splint: 5'-CACGCA TTTGAAACGTGGCCATAACTTTAGACATACCTTTCTAGATCACTTATTC-3' (ligation splint); T7 UBC4 37-135 F: 5'-TAATACGACTCACTATAGGCCACG TTTCAAAATGC-3' (forward primer to generate the template UBC4 TX); HindIII UBC4 R: 5'-ATAAGCTTAACATGAAGTAGGTGGATCTC-3' (reverse primer to generate the template for UBC4 TX).

The following oligonucleotides were used for synthesis of ACT1 substrates. ACT1-3'O-pc: 5'-UUUA(2'-o-nitrobenzyl-G)AGGUUGCUGCUUU-3'; ACT1-3'S-pc: 5'-UUUA(3'S, 2'-o-nitrobenzyl-G)AGGUUGCUGCUUU-3'; ACT1-UAc-3'O: 5'-UUUACACGUUGCUGCUUU-3'; ACT1-UAc-3'S: 5'-UUUA(3'S-C)AG GUUGCUGCUUU-3'; ACT1-UAc-splint: 5'-GAACCGTTATCAATAACCAAA GCAGCAACGTGTAAACATATAATATAGCAACAAAA-3'; ACT1-splint: 5'-G AACCGTTATCAATAACCAAAAGCAGCAACCTCTAAACATATAATATAGCA ACAAAAA-3'; act1-3'-end-for-1: 5'-TAATACGACTCACTATAGGTTATTGAT AACGGTTATTG-3' (forward DNA primer to generate ACT1 392-590 transcrip- tion template); act1-3'-end-for-2: 5'-GAAATTAATACGACTCACTATAGGTT ATTG-3' (forward DNA primer to generate ACT1 392-590 transcription tem- plate); act1-3'-end-rev: 5'-mUmUGGGCTGCAGGTGCGAGCTC-3' (reverse DNA primer to generate ACT1 392-590 transcription template); EcoRI+T7:5'-AGTG AATTCCTTAATACGACTCACTATAGG-3' (forward DNA primer to gener- ate ACT1 1-373-HDV template); Forward long: 5'-CGTAATACGACTCACT ATAGGGCGAATTGG-3' (forward DNA primer to generate ACT1 1-373-HDV template); Reverse-A-25: 5'-CGAGGAGGCTGGGAGCATGCCGGCCATAT AATATAGCAACAAAAAGAAT-3' (reverse DNA primer to generate ACT1 1-373-HDV template); Reverse 0-50: 5'-CCGGAATGTTGCCAGCCGGCGC CGCGAGGAGGCTGGGAGCATGCCGGCC-3' (reverse DNA primer to generate ACT1 1-373-HDV template); Reverse 25-75: 5'-GTGCGTCCCATTGCGCATT ACCGGACGGTCCGGAATGTTGCCAGCCGGCGCGC-3' (reverse DNA primer to generate ACT1 1-373-HDV template); Reverse final: 5'-GTGCGTCCCATTGCG CCATTACCCG-3' (reverse DNA primer to generate ACT1 1-373-HDV template).

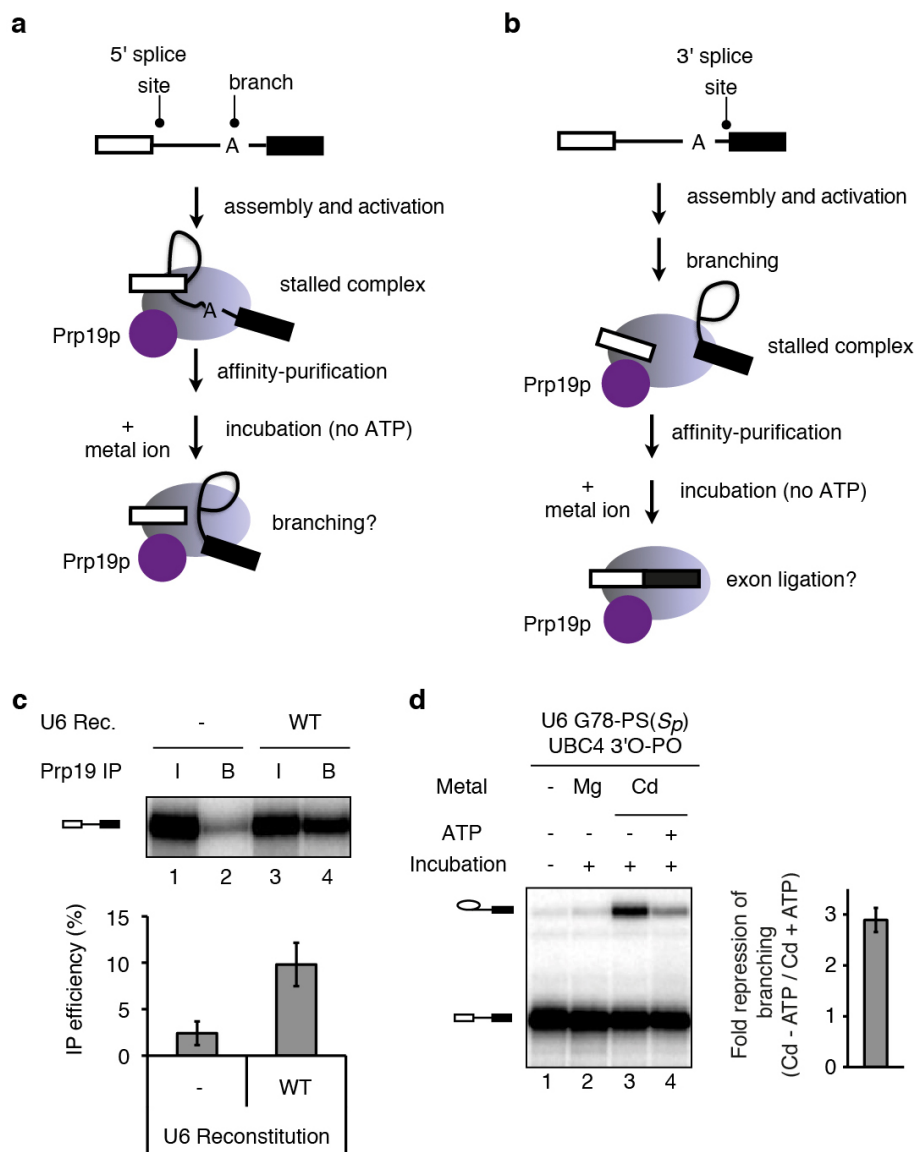
**Data analysis.** Gels were dried, exposed to storage phosphor screens (Amersham Biosciences) for 24 to 48 h, and scanned using a Typhoon Trio phosphorimager (Amersham Biosciences). Bands were quantified using ImageQuant TL with an automated rolling ball algorithm for background subtraction. The efficiency of branching was calculated as LI/(LI + P), if EI represented less than 1% of all species or (LI + EI)/(LI + EI + P) in all other cases; LI, lariat intermediate; EI,

excised intron; P, pre-mRNA. The efficiency of exon ligation was calculated as EI/(EI + LI).

The rescue midpoint for the metal titration curves in Fig. 3d and Extended Data Fig. 6c,d was obtained by fitting the rescue profiles to the general Hill equation  $A = A_{\max}x^n/(x^n + K^n)$ , where  $A_{\max}$  is the splicing efficiency at saturation,  $K$  is the rescue midpoint,  $x$  is the CdCl<sub>2</sub> concentration, and  $n$  is the Hill coefficient. Curves in Fig. 3e and Extended Data Fig. 7a were fit to the following equations:  $y = E_0 + Ex/(x + K_{Cd})$  (1 metal model) or  $y = E_0 + Ex^2/(x^2 + K_{Cd}^2)$  (2 metal model), where  $y$  is the branching efficiency,  $E_0$  is the extent of branching in the absence of Cd<sup>2+</sup>,  $E$  is the extent of branching at saturating Cd<sup>2+</sup>, and  $K_{Cd}$  is the apparent transition midpoint for Cd<sup>2+</sup> binding. Fits using the general Hill equation:  $y = E_0 + Ex^n/(x^n + K_{Cd}^n)$ , where  $n$  is the number of metal sites titrated, gave the same number of sites (within fit error) as those assuming a fixed  $n$  (data not shown).

Initial rates in Extended Data Figs 4d and 6a, b were obtained by fitting the linear portion of the splicing time courses to the equation  $A = A_0 + k_{\text{initial}}t$ , where  $A$  is the splicing efficiency at time  $t$  and  $A_0$  is the extent of splicing at time 0. Rates in Extended Data Fig. 9 were obtained by fitting the splicing time courses to the equation  $E = E_0 + A(1 - e^{-kt})$  where  $E$  is the extent of reaction,  $E_0$  is the extent at time = 0,  $A$  is the amplitude,  $k$  is the rate of reaction, and  $t$  is the time in minutes.

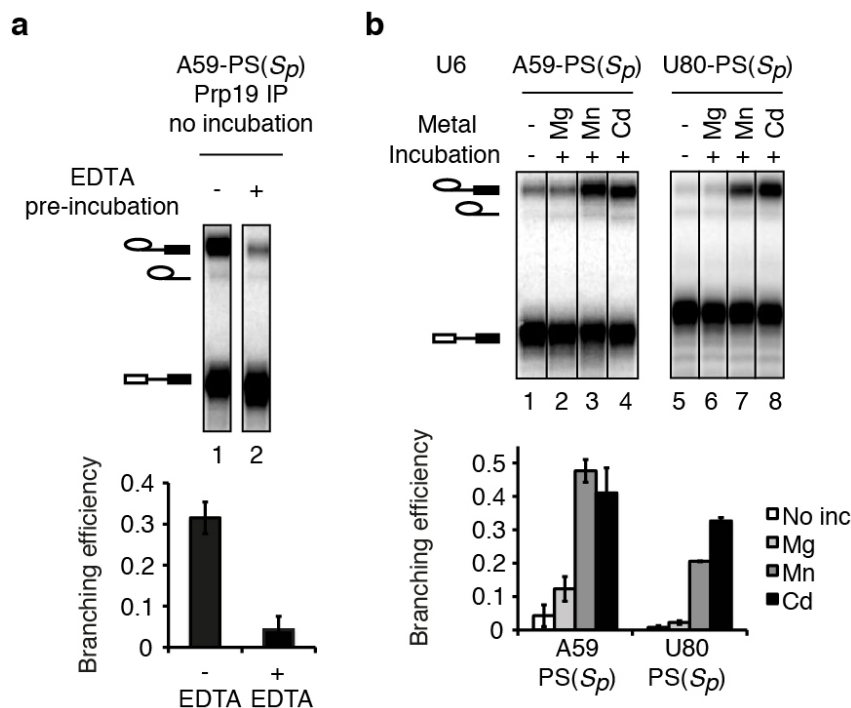
- Tagwerker, C. *et al.* HB tag modules for PCR-based gene tagging and tandem affinity purification in *Saccharomyces cerevisiae*. *Yeast* **23**, 623–632 (2006).
- Umen, J. G. & Guthrie, C. A novel role for a U5 snRNP protein in 3' splice site selection. *Genes Dev.* **9**, 855–868 (1995).
- Abelson, J., Hadjivassiliou, H. & Guthrie, C. Preparation of fluorescent pre-mRNA substrates for an smFRET study of pre-mRNA splicing in yeast. *Methods Enzymol.* **472**, 31–40 (2010).
- Frederiksen, J. K. & Piccirilli, J. A. Separation of RNA phosphorothioate oligonucleotides by HPLC. *Methods Enzymol.* **468**, 289–309 (2009).
- Loverix, S., Winqvist, A., Strömberg, R. & Steyaert, J. Mechanism of RNase T1: concerted triester-like phosphoryl transfer via a catalytic three-centered hydrogen bond. *Chem. Biol.* **7**, 651–658 (2000).
- Schürer, H., Lang, K., Schuster, J. & Mörl, M. A universal method to produce *in vitro* transcripts with homogeneous 3' ends. *Nucleic Acids Res.* **30**, e56 (2002).
- Dery, K. J., Yean, S.-L. & Lin, R.-J. Assembly and glycerol gradient isolation of yeast spliceosomes containing transcribed or synthetic U6 snRNA. *Methods Mol. Biol.* **488**, 41–63 (2008).
- Fabrizio, P., McPheeters, D. S. & Abelson, J. *In vitro* assembly of yeast U6 snRNP: a functional assay. *Genes Dev.* **3**, 2137–2150 (1989).
- Chiu, Y.-F. *et al.* Cwc25 is a novel splicing factor required after Prp2 and Yju2 to facilitate the first catalytic reaction. *Mol. Cell. Biol.* **29**, 5671–5678 (2009).
- Warkocki, Z. *et al.* Reconstitution of both steps of *Saccharomyces cerevisiae* splicing with purified spliceosomal components. *Nature Struct. Mol. Biol.* **16**, 1237–1243 (2009).
- Sigel, R. K. O., Song, B. & Sigel, H. Stabilities and structures of metal ion complexes of adenosine 5'-O-thiomonophosphate (AMPS<sup>2-</sup>) in comparison with those of its parent nucleotide (AMP<sup>2-</sup>) in aqueous solution. *J. Am. Chem. Soc.* **119**, 744–755 (1997).



**Extended Data Figure 1 | Affinity purification of spliceosomes with Prp19 requires reconstitution with U6 snRNA and enhances the potential to detect rescue by thiophilic metals.** **a, b,** Schemes depicting experimental strategy for staging spliceosomes to monitor branching (**a**) or exon ligation (**b**) in the absence of ATP. Spliceosomes are depicted as light magenta ovals and Prp19p as a magenta circle. Following affinity purification of spliceosomes by Prp19p (ref. 36), beads were washed to remove ATP and soluble factors and metals ions were added to assay for splicing. **c,** Prp19p-mediated affinity purification of activated spliceosomes, reflected by immunoprecipitation of pre-mRNA, is specific for properly reconstituted complexes. Note that the affinity

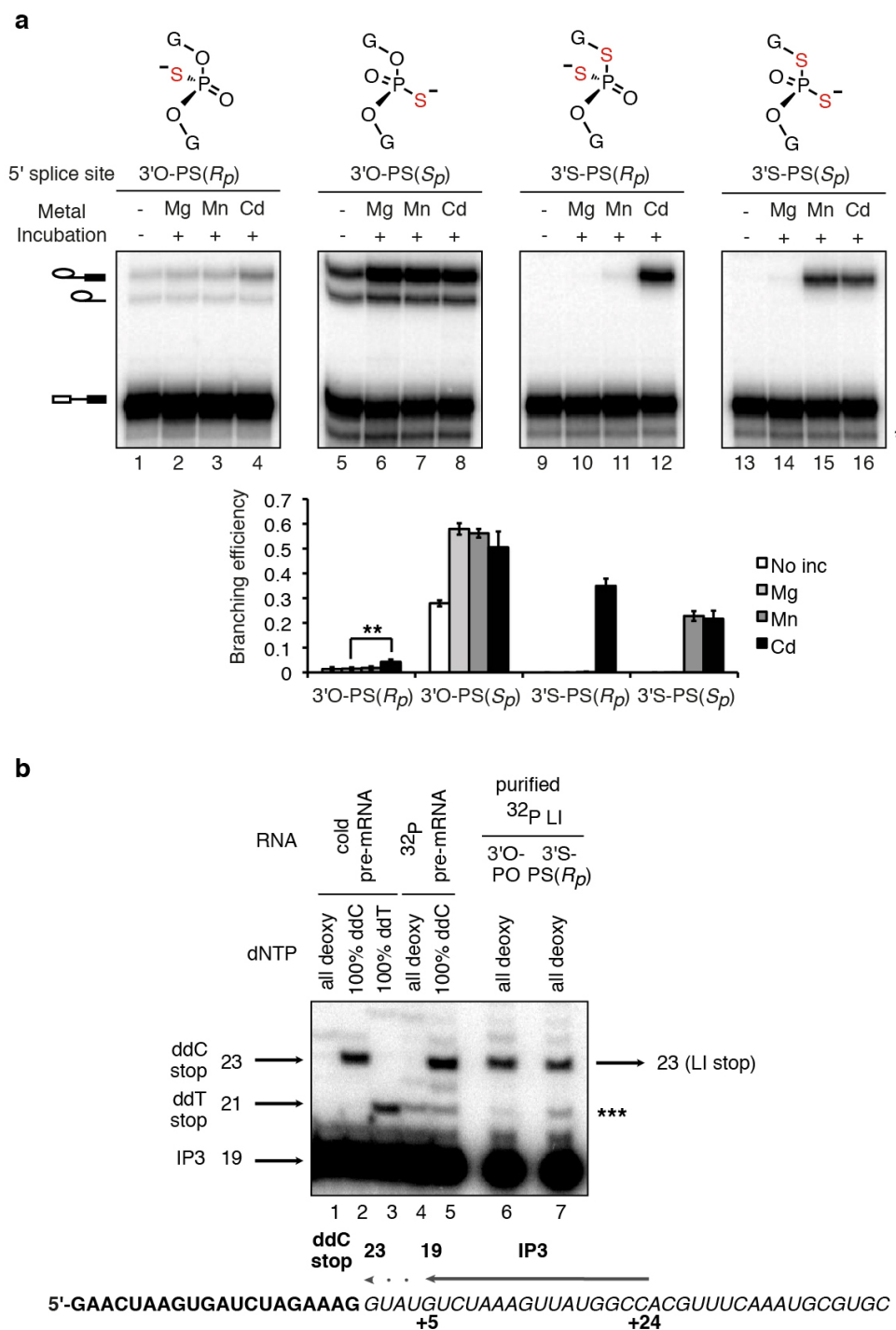
purification allows quantification of the branching efficiency for activated complexes, independently of any effects on assembly. RNA from 10% of the reaction (input, I) or from beads after affinity purification (B) was extracted and analysed by denaturing PAGE. Raw data (top); quantification of immunoprecipitation efficiency (bottom); Rec., reconstitution. **d,** ATP represses the Cd<sup>2+</sup> rescue for G78-PS(*S<sub>p</sub>*) spliceosomes. Spliceosomes were assayed as in Fig. 2e; for lane 4, 2 mM ATP-Mg<sup>2+</sup> was also present during the incubation. Representative gel (left) and quantification of the extent of ATP repression (right). Values are averages; error bars represent s.d. (*n* = 3).





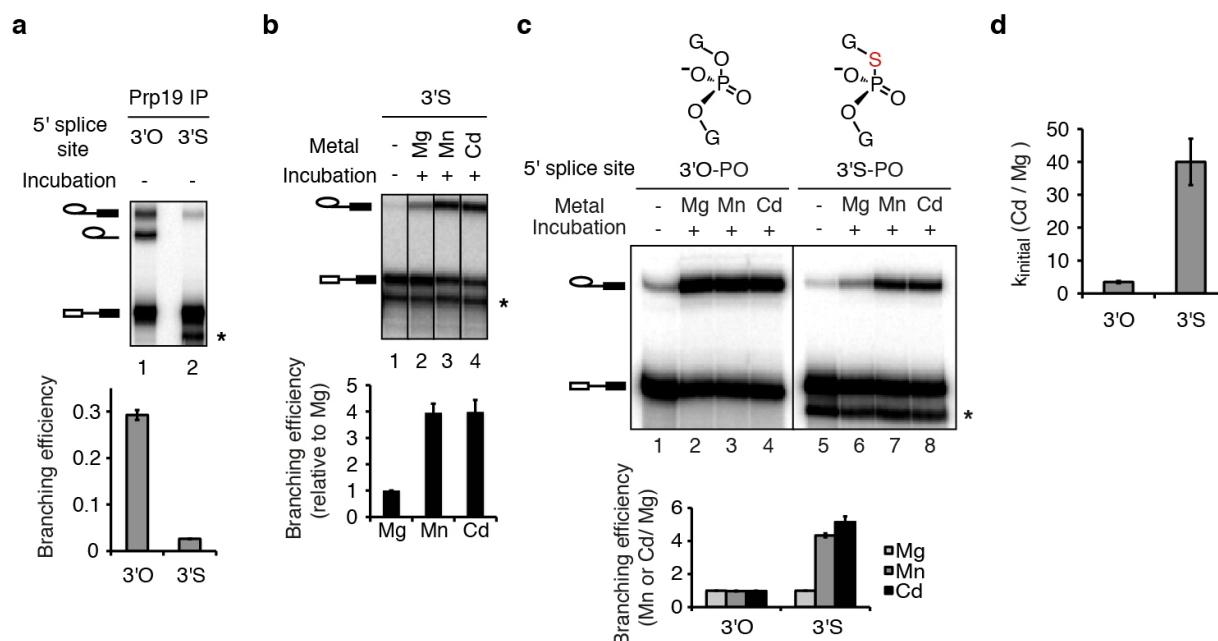
**Extended Data Figure 2 | Broad rescue specificity of A59-PS( $S_p$ ) and U80-PS( $S_p$ ) spliceosomes.** **a**, Pre-incubation with EDTA reveals a branching defect for A59-PS( $S_p$ ). Extracts were pre-incubated at 4 °C in the presence or absence of 2 mM EDTA and then incubated with the 3' O substrate. After affinity-purification, branching efficiency was quantified without further incubation (bottom). A representative gel is shown (top). EDTA pre-incubation caused an eightfold reduction in branching efficiency, indicating that splicing extracts may contain a thiophilic metal that supports branching by A59-PS( $S_p$ ) spliceosomes; note that pre-mRNA still immunoprecipitated efficiently, indicating catalytic activation of the spliceosome. In contrast EDTA pre-incubation has no effect on U6 wild-type spliceosomes (data not shown). **b**, The

branching defects for A59-PS( $S_p$ ) and U80-PS( $S_p$ ) spliceosomes are rescued by either  $Mn^{2+}$  or  $Cd^{2+}$ . Assays were as in Fig. 2c. A representative gel (top) and quantification (bottom) are shown; no inc, no incubation. Both  $Mn^{2+}$  and  $Cd^{2+}$  strongly rescue A59-PS( $S_p$ ) and U80-PS( $S_p$ ) spliceosomes (lanes 3, 4, 7 and 8), suggesting that even the weaker  $Mn^{2+}$ -S interaction<sup>38</sup> at these positions can support branching. This broad specificity for branching may also explain why A59-PS( $S_p$ ) and U80-PS( $S_p$ ) spliceosomes also catalysed exon ligation in the presence of thiophilic metals, whereas G60-PS( $R_p$ ), G78-PS( $S_p$ ), and U80-PS( $R_p$ ) spliceosomes, for which branching was only rescued in the presence of  $Cd^{2+}$ , stalled after branching. Values are averages; error bars, s.d. ( $n = 3$ ).



**Extended Data Figure 3 | A divalent metal binds the 5' splice site *pro-R<sub>p</sub>* oxygen during branching.** **a**, *UBC4* pre-mRNAs bearing the indicated modifications at the 5' splice site (top panel) were assayed as in Fig. 3a. The band marked \* results from 5' to 3' exonucleolytic degradation that is blocked by the sulphur; \*\* denotes statistical significance of Cd<sup>2+</sup> rescue compared to Mg<sup>2+</sup> splicing ( $P = 0.004$ , paired, one-tailed  $t$ -test,  $n = 3$ ). The data from Fig. 3a are reproduced here to aid comparison. Values are averages; error bars, s.d. ( $n = 3$ ); no inc, no incubation. **b**, Mapping of the site of branching. Purified, <sup>32</sup>P-labelled, intact lariat intermediates (LI ~ 0.2 fmol) resulting from branching of the *UBC4* 3'O-PO or 3'S-PS(*R<sub>p</sub>*) substrates were used as templates for reverse transcriptase (RT) using primer IP3 (~10 fmol), which

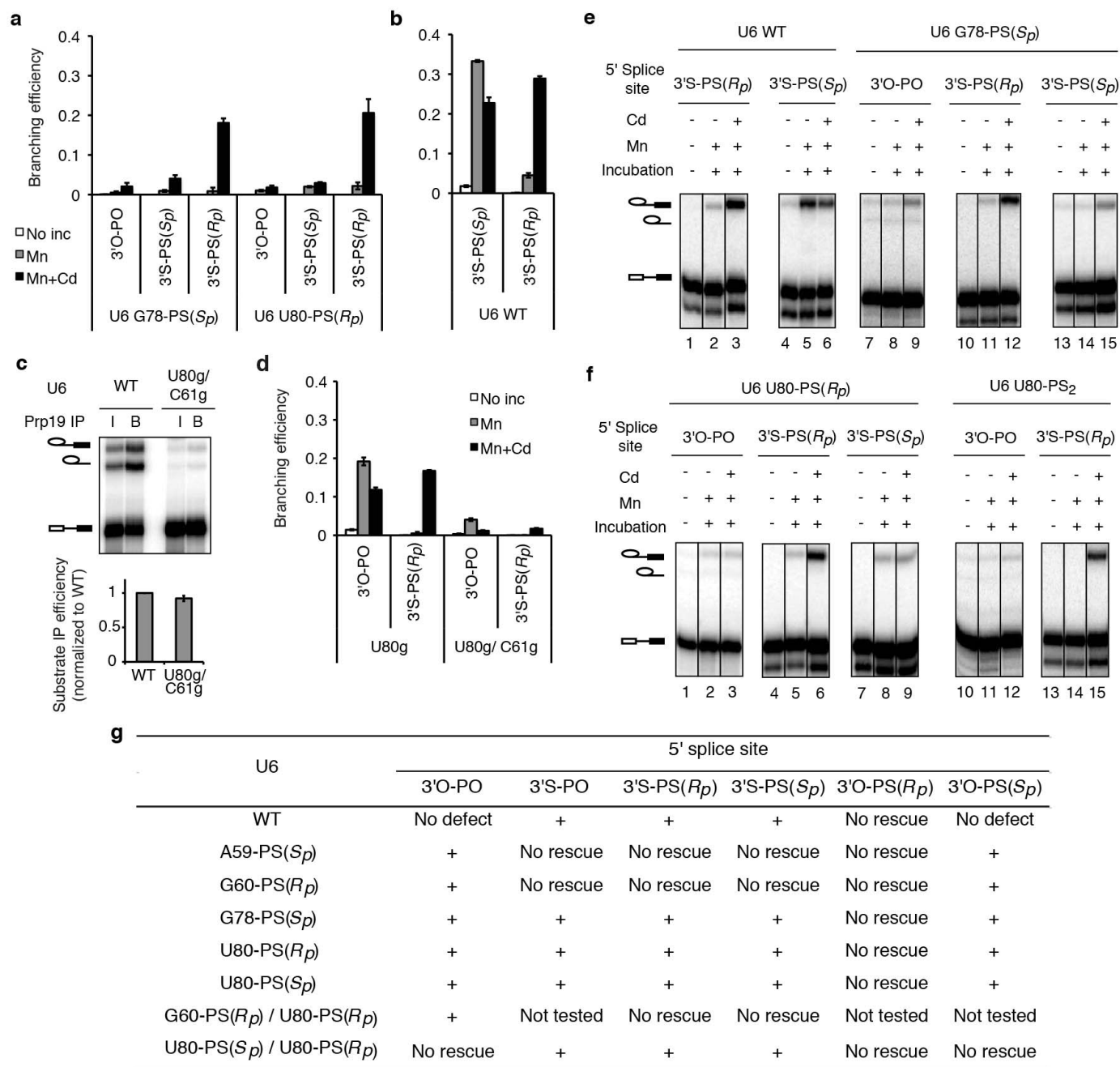
binds at nucleotides +5 to +24 of the intron (see lower diagram). Lanes 6 and 7 show that the major RT stop occurs at the same position when either the 3'O-PO or 3'S-PS(*R<sub>p</sub>*) lariats are used as template. This stop migrates at the expected position, which is the position of the ddC stop resulting from extension of primer IP3 with pre-mRNA as template and therefore corresponds to position +1 of the intron, the expected branch site. Lower diagram shows mapping of the primer and expected RT stop onto the *UBC4* pre-mRNA sequence; bold exon; italics intron. Note that the band marked \*\*\* is present in the <sup>32</sup>P pre-mRNA lane and thus is probably a nonspecific band resulting from contamination with pre-mRNA degradation products that can anneal to the primer and serve as template.



**Extended Data Figure 4 | The *UBC4* 3'S substrate can be branched efficiently only in the presence of thiophilic metals.** **a**, A sulphur at the 5' splice site leaving group alone blocked splicing in extract. Following affinity-purification branching efficiency was quantified (bottom) without further incubation. Note that immunoprecipitation of pre-mRNA indicated that the sulphur did not block the catalytic activation of the spliceosome. **b**, Thiophilic metals rescued the branching defect of the 3'S substrate. Spliceosomes were assayed in buffer PK (pH 7.0) in the presence of 2 mM total metal (1 mM MgCl<sub>2</sub> plus 1 mM MnCl<sub>2</sub> or 1 mM CdCl<sub>2</sub>) for 1.5 h. The bar graph quantifies the relative stimulation by specific metals (normalized to Mg<sup>2+</sup>). **c**, **d**, Thiophilic

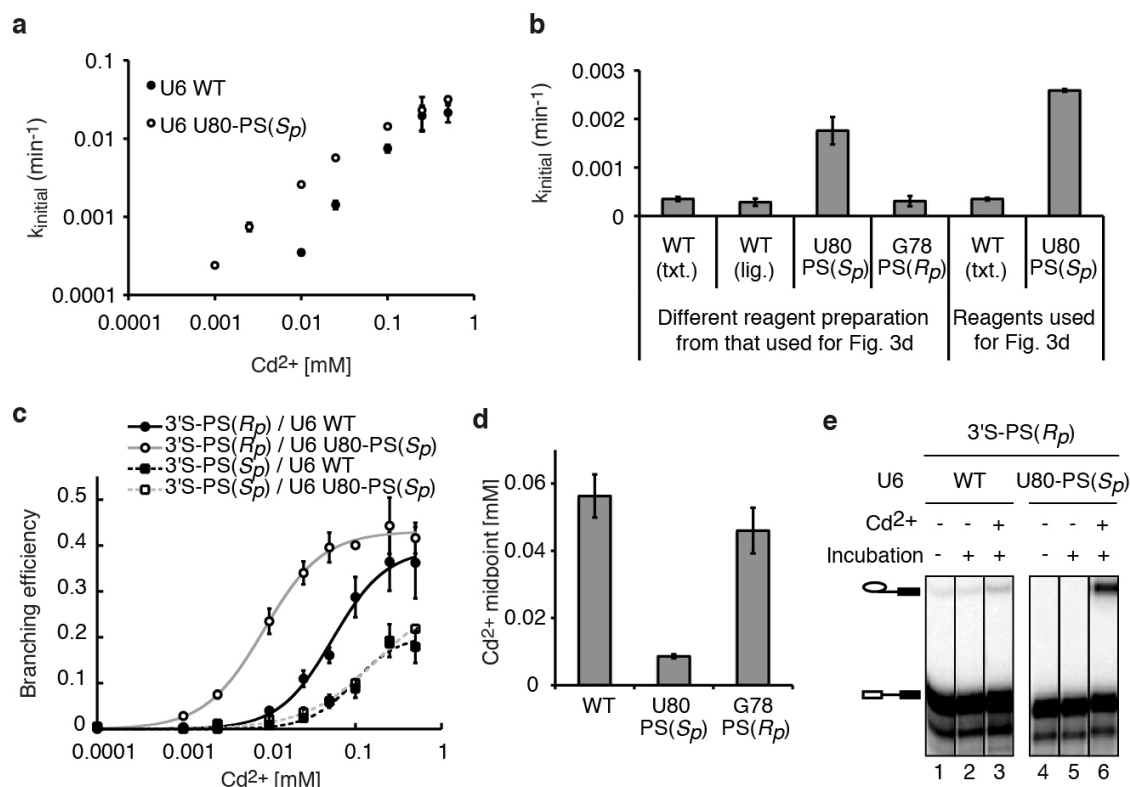
metals specifically stimulate branching of the 3'S, but not the 3'O substrate. Affinity-purified spliceosomes from extracts depleted of Cwc25p, to stall spliceosomes independent of the sulphur substitution, were incubated as in Fig. 3a except that washes and incubation were done in buffer DK without EDTA and with 1 mM MgCl<sub>2</sub>; rCwc25 as well as an HP extract fraction were also added to complement (Supplementary Note 4; Methods). Quantification of the thiophilic metal stimulation relative to Mg<sup>2+</sup> is shown both for reaction endpoints (**c**) and for the rate of branching (**d**). Values are averages, s.d. ( $n = 3$ ). The band marked \* results from 5' to 3' exonucleolytic degradation that is blocked by the sulphur.





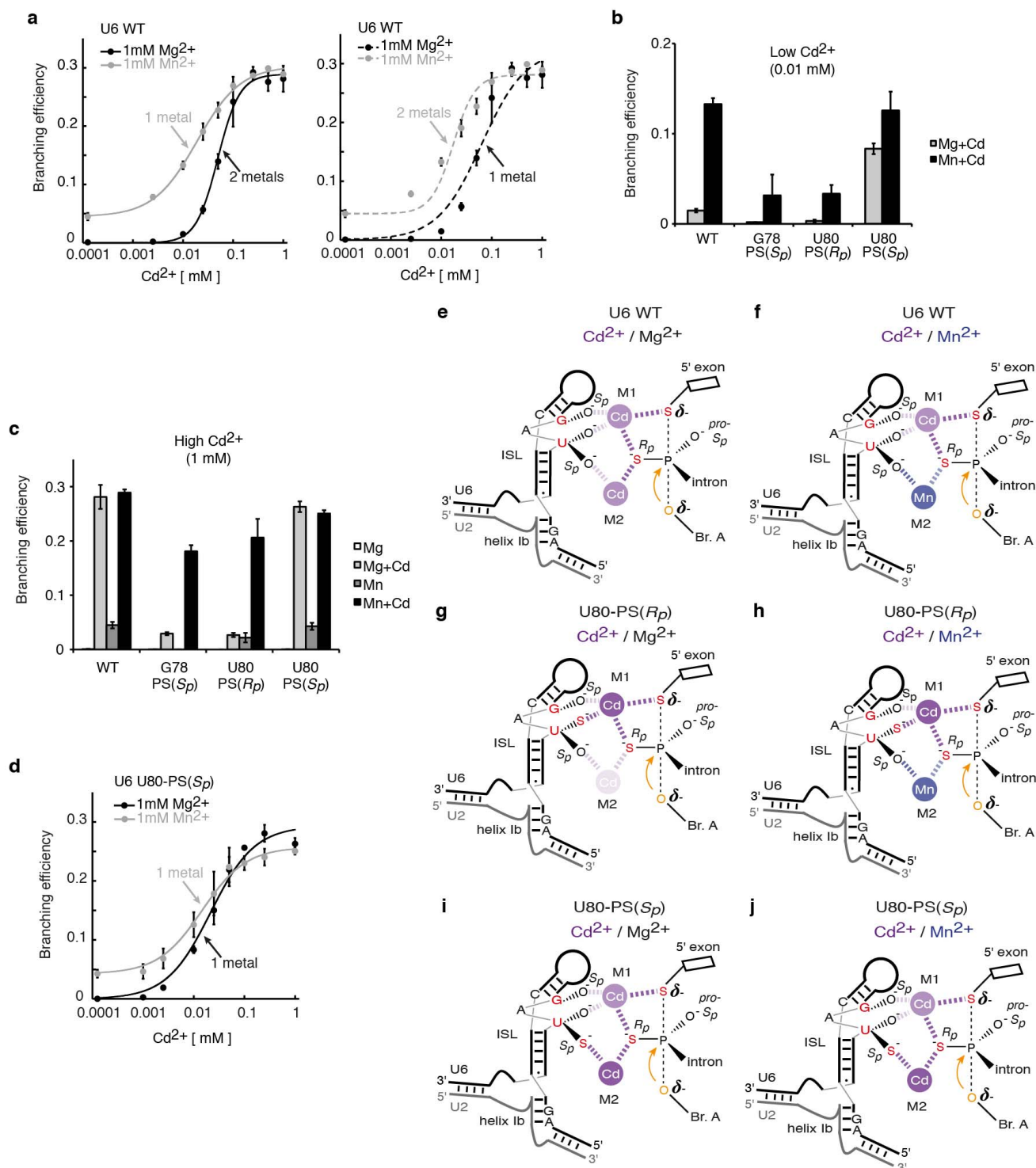
**Extended Data Figure 5 | The 3'S-PS(*R<sub>p</sub>*) substrate specifically improves rescue for spliceosomes containing U6 sulphur substitutions that compromise catalytic metal binding.** **a, b**, Spliceosomes were assayed as in Fig. 3b. Note that the 3'S-PS(*S<sub>p</sub>*) substrate does not significantly improve Cd<sup>2+</sup> rescue when compared to the 3'O-PO substrate (**a**), despite having similar reactivity to the 3'S-PS(*R<sub>p</sub>*) substrate with wild-type U6 (**b**). **c**, The U6 double mutation U80g/C61g permitted both spliceosome assembly and activation, as reflected by the stable association of Prp19p with the splicing substrate. 10% of the RNA in the input (I) for the immunoprecipitation or 100% of the RNA associated with affinity-purified spliceosomes (B) were analysed by denaturing PAGE (top). Total immunoprecipitation efficiency was quantitated for all

splicing species combined (bottom). **d**, The 3'S-PS(*R<sub>p</sub>*) substrate did not significantly improve splicing for U80g and U80g/C61g spliceosomes. Assays were as in Fig. 3b. **e, f**, Representative raw data for Fig. 3. Assays were as in Fig. 3. In **e**, for wild-type U6 lanes 1–3 and 4–6 were taken from two different gels, for G78-PS(*S<sub>p</sub>*) lanes 7–12 and 13–15 were taken from two different gels. In all other cases the lanes for different substrates assembled with spliceosomes bearing the same U6 modification were taken from the same gel. Values are averages; error bars, s.d. ( $n = 3$  for **a, b, d**;  $n = 2$  for **c**); no inc., no incubation. **g**, Summary of combinations of sulphur substitutions in U6 and the substrate tested for rescue of branching. The + sign indicates that branching was observed in the presence of thiophilic metal.



**Extended Data Figure 6 | U6 snRNA positions catalytic metals during branching: controls for the U80-PS( $S_p$ ) induced shift in the  $\text{Cd}^{2+}$  transition midpoint for rescue of the 3'S-PS( $R_p$ ) substrate.** **a**, The shift induced by U80-PS( $S_p$ ) in the  $\text{Cd}^{2+}$  midpoint for rescue was also observed when reaction rates, instead of amplitudes, were compared. Initial rates are plotted versus  $\text{CdCl}_2$  concentration. Assays were as in Fig. 3d. Values are averages; error bars represent s.d. ( $n = 2$ ). Initial rates, rather than apparent overall rates, were used here because the branching efficiency did not level off by 120 min at  $\text{Cd}^{2+}$  concentrations below 0.025 mM (assuming an endpoint of  $\sim 0.4$ ). In support of this approach, at saturating  $\text{Cd}^{2+}$  both wild-type and U80-PS( $S_p$ ) spliceosomes branched similar fractions of the 3'S-PS( $R_p$ ) substrate (see **c**), indicating that addition of a sulphur in U6 within the catalytic core did not necessarily alter the population of complexes that are competent for catalysis. **b**, Transcribed (txt.) and ligated (lig.) wild-type U6 behaved similarly relative to U80-PS( $S_p$ ) for branching of the 3'S-PS( $R_p$ ) substrate at a limiting  $\text{Cd}^{2+}$  concentration. Initial rates are shown for branching of the 3'S-PS( $R_p$ ) substrate in the presence of 0.01 mM  $\text{Cd}^{2+}$ . **c**, A sulphur at U80 *pro-S<sub>p</sub>* shifts the  $\text{Cd}^{2+}$  titration midpoint

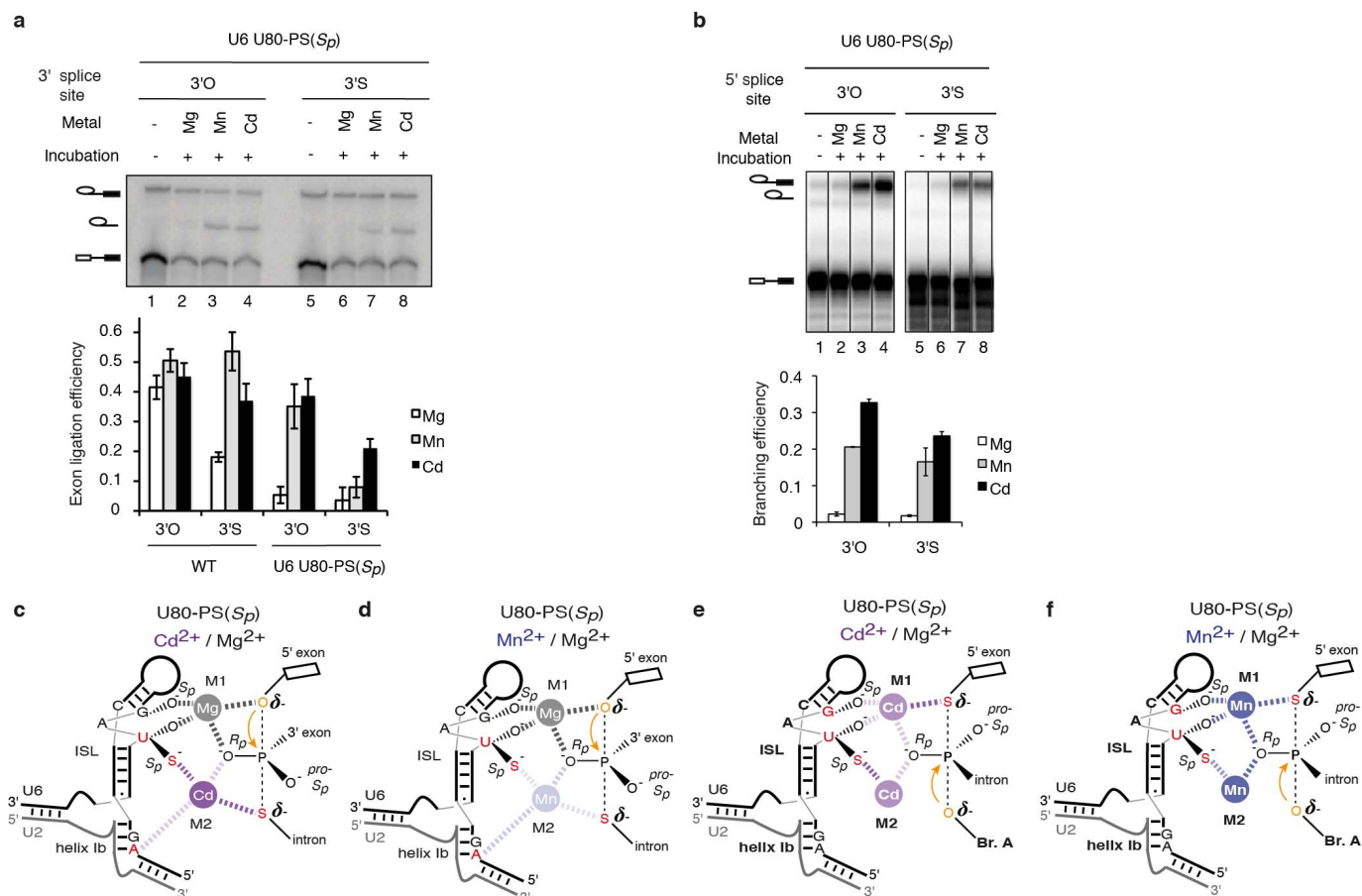
for rescue of the 3'S-PS( $R_p$ ) substrate relative to the 3'S-PS( $S_p$ ) substrate. Assays were as in Fig. 3d; the 3'S-PS( $R_p$ ) data from Fig. 3d are shown again here for comparison. Curves represent Hill fits to the data. Values are averages; error bars, s.d. ( $n = 3$ ). Although the  $\text{Cd}^{2+}$  titration for rescue of U80-PS( $S_p$ ) spliceosomes assembled on the 3'S-PS( $S_p$ ) substrate did not plateau under our experimental conditions (panel **c**), the data nevertheless set a lower limit for the apparent transition midpoint; further, this midpoint is equal to or greater than that observed for U6 wild-type spliceosomes, indicating that the shift by U80-PS( $S_p$ ) of the  $\text{Cd}^{2+}$  transition midpoint for rescue was specific for the 3'S-PS( $R_p$ ) substrate. **d**, With the 3'S-PS( $R_p$ ) substrate, U80-PS( $S_p$ ) decreased the  $\text{Cd}^{2+}$  titration midpoint for rescue of branching by sixfold. The apparent midpoint for G78-PS( $R_p$ ) spliceosomes is shown as an additional specificity control for the shift observed for U80-PS( $S_p$ ) (actual titration data not shown). The apparent  $\text{Cd}^{2+}$  rescue midpoints were obtained by fitting titration curves to the general Hill equation (see Methods). Error bars represent error of the Hill fit. **e**, Representative raw data for Fig. 3d. Assays were in the presence of 0.01 mM  $\text{Cd}^{2+}$  when indicated.



**Extended Data Figure 7 | Further evidence for two distinct catalytic metal sites during branching.** **a**, Branching requires two catalytic divalent metals. **b–d**, The 5' splice site *pro-R<sub>p</sub>* oxygen and the U80 *pro-S<sub>p</sub>* oxygen interact with a metal distinct from the metal that interacts with the 5' splice site *pro-R<sub>p</sub>* oxygen and the U80 *pro-R<sub>p</sub>* and G78 *pro-S<sub>p</sub>* oxygens. See Supplementary Note 7 for discussion. Spliceosomes were assayed for branching as in Fig. 3e; where indicated,  $\text{MgCl}_2$  or  $\text{MnCl}_2$  were present at 1 mM. Curves represent Hill fits to the data. The data in **a** are reproduced here from Fig. 3e to aid comparison. In panels **a–d**, values are averages; error bars, s.d. ( $n = 3$ ).  $\text{Cd}^{2+}$  was limiting in **b** to sensitize the assay to binding of a second metal, and  $\text{Cd}^{2+}$  was saturating in **c** to show that G78-PS( $S_p$ ) and U80-PS( $S_p$ ) have the potential to be rescued at levels comparable to wild-type U6. Panel **d** shows that U80-PS( $S_p$ ) eliminates

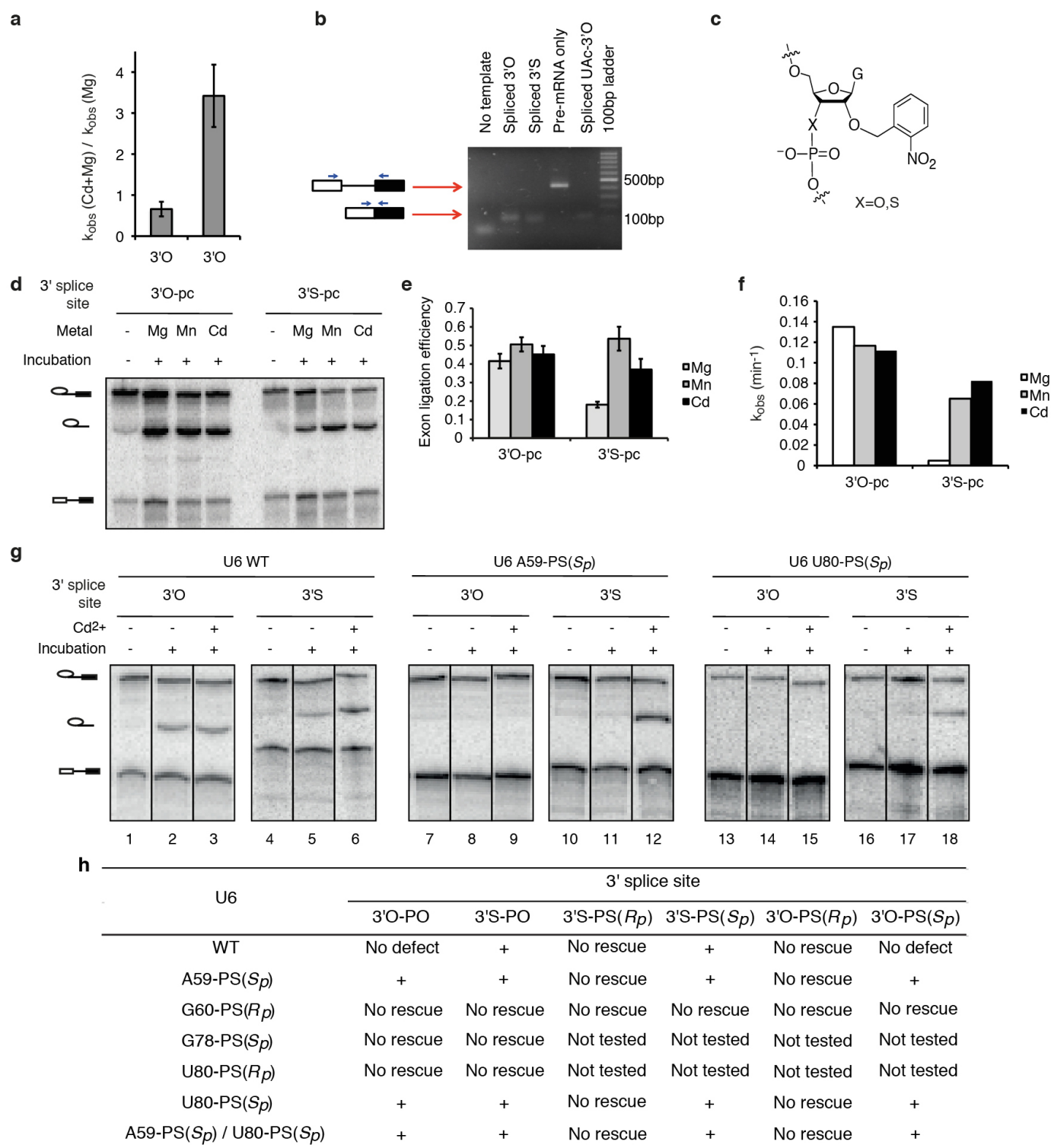
the affect of  $\text{Mn}^{2+}$  on the titration curves. **e–j**, Metal binding during branching for different combinations of sulphur substitutions in the substrate and U6 in the presence of the indicated metals, as reflected by the data in panels **a–d**. Panels **e** and **f** reflect data in **a**, **b** and **c**; panels **g** and **h** reflect data in **b** and **c**; panels **i** and **j** reflect data in **b**, **c** and **d**. Relevant U6 ligands are coloured red in each panel; the nucleophile is coloured orange. Metals are coloured magenta ( $\text{Cd}^{2+}$ ) and blue ( $\text{Mn}^{2+}$ ), and their interactions with specific U6 ligands are depicted as dashed lines, with differential shading intensity meant to illustrate differences in the expected strength of interaction with oxygen compared with sulphur, as inferred from studies with model compounds<sup>38,61</sup>. Shading of metals bound at M1 and M2 is further adjusted to reflect experimental observations. Panels for G78-PS( $S_p$ ) would look the same as those for U80-PS( $R_p$ ) (**g** and **h**).





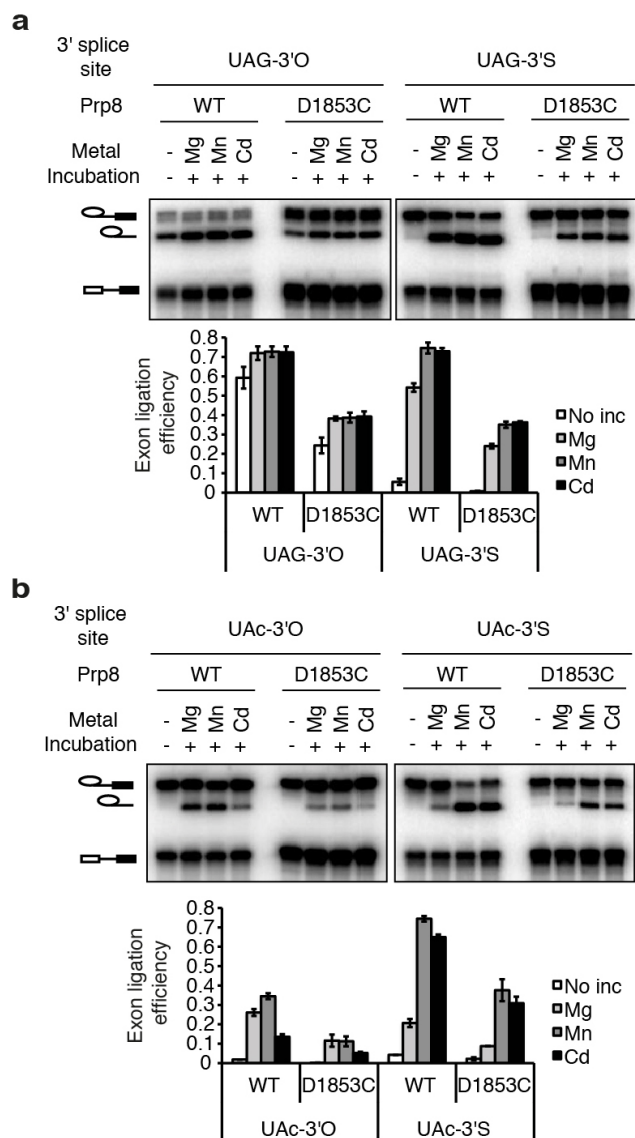
**Extended Data Figure 8 | A sulphur at the 3' splice site but not the 5' splice site leaving group alters the metal specificity for rescue of U80-PS( $S_p$ ) spliceosomes: further evidence that U80 interacts with a catalytic metal during exon ligation.** **a**, A sulphur at the 3' splice site leaving group during exon ligation alters the metal specificity for rescue of U80-PS( $S_p$ ) spliceosomes. Assays were as in Fig. 2j. The bar graph depicts quantification of exon ligation efficiency. **b**, A sulphur at the 5' splice site leaving group during branching does not alter the metal specificity for rescue of U80-PS( $S_p$ )

spliceosomes. Assays were as in Fig. 3a, except with the 3'S-PO substrate. Values are averages; error bars, s.d. ( $n = 3$ ). **c–f**, Inferred metal binding during exon ligation (**c**, **d**) and branching (**e**, **f**) for the indicated combinations of sulphur substitutions in the substrate and U6 in the presence of the indicated metals. Relevant U6 ligands are coloured red in each panel; the nucleophile is coloured orange. Metals and ligand interactions are coloured as in Extended Data Fig. 7.



**Extended Data Figure 9 | Thiophilic metals rescue exon ligation for substrates containing a sulphur at the 3' splice site leaving group in both a mutated and wild-type 3' splice site context.** **a**,  $\text{Cd}^{2+}$  specifically stimulates the rate of exon ligation for the mutated, UAc-3'S substrate. Assays were as in Fig. 4a. **b**, Exon ligation occurs at the correct site for the 3'S substrate. RNA from affinity-purified spliceosomes assembled on ACT1-3'O, ACT1-3'S, and ACT1-UAc-3'O and chased as in Fig. 4a was subjected to RT-PCR using the primers depicted in blue arrows. **c**, Diagram of the photocaged linkage at the 3' splice site. **d-f**,  $\text{Cd}^{2+}$  specifically stimulates the rate of exon ligation for the

wild-type, UAG-3'S substrate. Assays were as in Fig. 4a, except that before addition of divalent metals, samples were irradiated with 308 nm light on ice for 5 min to remove the photocage. Shown are a representative gel (**d**), quantification of the reaction end points (**e**) and quantification of reaction rates (**f**). **g**, Representative raw data for Fig. 4b; bands within each set came from non-adjacent lanes on the same gel. Values are averages; error bars, s.d. ( $n = 3$ ). **h**, Summary of combinations of sulphur substitutions in U6 and the substrate tested for rescue of exon ligation. The + sign indicates that exon ligation was observed in the presence of thiophilic metal.



**Extended Data Figure 10 | Residue D1853 of the RNaseH-like domain of Prp8 does not have a direct role in metal-mediated catalysis of exon ligation.**

**a, b,** Spliceosomes assembled on the indicated ACT1 UAG (**a**) or UAc (**b**) 3'O or 3'S substrates were assayed as in Fig. 2j, in the absence of ATP. Splicing extracts were prepared from either a wild-type *PRP8* strain or a mutant strain having the *prp8-D1853C* mutation. See Supplementary Note 15 for details and a discussion. Values are averages; error bars, s.d. ( $n = 2$ ); no inc, no incubation.



# The trajectory, structure and origin of the Chelyabinsk asteroidal impactor

Jiří Borovička<sup>1</sup>, Pavel Spurný<sup>1</sup>, Peter Brown<sup>2,3</sup>, Paul Wiegert<sup>2,3</sup>, Pavel Kalenda<sup>4</sup>, David Clark<sup>2,3</sup> & Lukáš Šrbený<sup>1</sup>

Earth is continuously colliding with fragments of asteroids and comets of various sizes. The largest encounter in historical times occurred over the Tunguska river in Siberia in 1908, producing<sup>1,2</sup> an airburst of energy equivalent to 5–15 megatons of trinitrotoluene (1 kiloton of trinitrotoluene represents an energy of  $4.185 \times 10^{12}$  joules). Until recently, the next most energetic airburst events occurred over Indonesia<sup>3</sup> in 2009 and near the Marshall Islands<sup>4</sup> in 1994, both with energies of several tens of kilotons. Here we report an analysis of selected video records of the Chelyabinsk superbolide<sup>5</sup> of 15 February 2013, with energy equivalent to 500 kilotons of trinitrotoluene, and details of its atmospheric passage. We found that its orbit was similar to the orbit of the two-kilometre-diameter asteroid 86039 (1999 NC43), to a degree of statistical significance sufficient to suggest that the two were once part of the same object. The bulk strength—the ability to resist breakage—of the Chelyabinsk asteroid, of about one megapascal, was similar to that of smaller meteoroids<sup>6</sup> and corresponds to a heavily fractured single stone. The asteroid broke into small pieces between the altitudes of 45 and 30 kilometres, preventing more-serious damage on the ground. The total mass of surviving fragments larger than 100 grams was lower than expected<sup>7</sup>.

The data for Tunguska are limited to tree damage and records of seismic or acoustic waves at large distances. The Indonesia and Marshall Islands impacts were detected only by distant infrasonic stations or satellites in orbit and were therefore poorly documented. Some camera data exist for the multikilaton Sutter's Mill event<sup>8</sup>, but precise analyses of the ablation process based on imaging have been made only for impacts of metre-sized bodies with energies of  $\sim 0.1$  kt trinitrotoluene<sup>9–11</sup> and less<sup>12</sup>.

The Chelyabinsk impact occurred unexpectedly over a relatively densely populated Russian region during sunrise on 15 February 2013. The superbolide (an extremely bright meteor) generated a damaging air blast wave. An 8-m-wide hole in the ice of Lake Chebarkul, 70 km west of Chelyabinsk, was reported shortly after the event. Thousands of small meteorites of total mass  $>100$  kg, classified as LL5 ordinary chondrites, were found in the areas south-southwest of Chelyabinsk<sup>13</sup>.

Here we determine the bolide trajectory and orbit and describe the ablation process of the asteroid. The main data for these analyses were 15 bolide videos publicly available on the internet (Extended Data Table 1). We calibrated these videos with wide-field stellar imagery. Details of our procedure, which was based on the least-squares method<sup>14</sup>, are given in Supplementary Information. The trajectory and speed of the bolide are presented in Table 1. The observed low deceleration provides an extreme lower limit of  $10^6$  kg for the mass of the body. The measured energy<sup>5</sup> and speed provide a best estimate of the mass of  $\sim 1.2 \times 10^7$  kg, corresponding to a diameter of  $\sim 19$  m assuming a bulk density of  $3,300 \text{ kg m}^{-3}$ .

The pre-impact orbit (Table 2) is consistent with an origin in the main asteroid belt, most probably in the inner main belt near the  $\nu_6$  secular resonance. We integrated the orbit and 1,000 test particles within the orbital uncertainties (a probability cloud) 2,000 years into the past. The asteroid spent the six weeks before impact within an elongation of  $45^\circ$

from the Sun, a region of the sky inaccessible to ground-based telescopes. At earlier times, the asteroid was always too faint to be seen when some portion of the probability cloud was in the field of view of existing asteroid surveys. We note that the 2.2-km-diameter<sup>15</sup> near-Earth asteroid 86039 (1999 NC43) of spectral type Q<sup>16</sup> (corresponding to ordinary chondrites) has a very similar orbit, with very low dissimilarity criteria,  $D = 0.050$  (ref. 17) and  $D' = 0.018$  (ref. 18), relative to Chelyabinsk asteroid. Though this does not provide an unequivocal dynamical link, such a close match is unlikely statistically. We expect 227 near-Earth asteroids brighter than 86039 to exist<sup>19</sup>. Selecting at random from the expected distribution of near-Earth asteroids<sup>20</sup>, it takes an average of  $6 \times 10^5$  draws before selecting one with a smaller  $D$  value, and more than  $3 \times 10^6$  draws before selecting one with a smaller  $D'$  value. Because 227/600,000 and 227/3,000,000 are equivalent to 1:2,600 and 1:13,000, respectively, we conclude that there is an approximately 1:10,000 chance that the proximity of these orbits is due purely to chance. The two orbits have maintained two intersection points over the past 2,000 years, one near perihelion and one near aphelion (Extended Data Fig. 1). The minimum velocity kick required to eject the Chelyabinsk asteroid from 86039 is  $0.7 \text{ km s}^{-1}$  (aphelion) or  $2 \text{ km s}^{-1}$  (perihelion). This ejection velocity is consistent with a collision with another asteroid (which would provide a kick of a few kilometres per second).

The fragmentation during atmospheric entry was studied using the bolide light curve<sup>5</sup> and deceleration with a procedure developed recently<sup>11</sup>. The arrival times of secondary sonic booms heard on videos were also used to locate fragmentation points (the primary blast wave was caused by the cylindrical shock from the trajectory<sup>5</sup>). The dynamic pressure

**Table 1 | Trajectory of the Chelyabinsk superbolide**

Time (s)	Longitude (°)	Latitude (°)	Height (km)	Speed (km s <sup>-1</sup> )
1.07	64.477	54.454	95.0	19.03
6.97	62.888	54.664	60.0	19.05
10.46	61.933	54.780	40.0	19.03
12.24	61.442	54.837	30.0	18.9
13.18	61.193	54.864	25.0	18.0
14.18	60.943	54.892	20.0	14.2
15.17	60.802	54.907	17.2	6
Fragment F1				
14.32	60.945	54.893	20.0	13.5
16.04	60.704	54.922	15.0	6.4
17.80	60.5883	54.9361	12.57	3.2

Time zero corresponds to approximately 3:20:20 UT. Coordinates are given in the WGS84 geoid system. Speed is relative to the Earth's surface. At the beginning, the acceleration due to gravity was larger than deceleration due to atmospheric drag. The beginning speed ( $19.03 \pm 0.13 \text{ km s}^{-1}$ ) remained constant to within  $0.02 \text{ km s}^{-1}$  down to an altitude of 35 km. The bolide was recorded in the analysed videos between altitudes of 95.1 and 12.6 km, over a trajectory of total length 272 km. The apparent radiant was changing owing to Earth's gravity, starting at right ascension  $\alpha = 21 \text{ h } 53 \text{ min } 7 \text{ s} \pm 24 \text{ s}$  ( $328.28^\circ \pm 0.10^\circ$ ) and declination  $\delta = +7^\circ 28' \pm 3'$  ( $+7.47^\circ \pm 0.05^\circ$ ) at altitude 95 km and changing to  $\alpha = 21 \text{ h } 51 \text{ min } 58 \text{ s}$  ( $327.99^\circ$ ),  $\delta = +7^\circ 53'$  ( $7.88^\circ$ ) at altitude 20 km. These values correspond to a local radiant azimuth of  $A = 103.50^\circ \pm 0.09^\circ$  (east of north) and elevation  $h = 18.55^\circ \pm 0.08^\circ$ , and, respectively,  $A = 100.63^\circ$  and  $h = 16.95^\circ$  (we note that  $A$  and  $h$  change also as a result of Earth's curvature). For comparison, the radiant of fragment F1 at 20 km was  $A = 101.87^\circ \pm 0.4^\circ$ ,  $h = 17.53^\circ \pm 0.3^\circ$ .

<sup>1</sup>Astronomical Institute, Academy of Sciences of the Czech Republic, CZ-251 65 Ondřejov, Czech Republic. <sup>2</sup>Department of Physics and Astronomy, University of Western Ontario, London, Ontario N6A 3K7, Canada. <sup>3</sup>Centre for Planetary Science and Exploration, University of Western Ontario, London, Ontario N6A 5B7, Canada. <sup>4</sup>Institute of Rock Structure and Mechanics, Academy of Sciences of the Czech Republic, V Holešovičkách 41, CZ-18209 Praha 8, Czech Republic.

**Table 2 | Pre-impact orbit of Chelyabinsk asteroid and the orbit of asteroid 86039 (1999 NC43)**

	Chelyabinsk	86039
Semimajor axis, $a$ (AU)	$1.72 \pm 0.02$	$1.759621064 \pm 6 \times 10^{-9}$
Perihelion distance, $q$ (AU)	$0.738 \pm 0.002$	$0.7403896 \pm 1 \times 10^{-7}$
Eccentricity, $e$	$0.571 \pm 0.006$	$0.57923346 \pm 6 \times 10^{-8}$
Inclination, $i$ (°)	$4.98 \pm 0.12$	$7.12312 \pm 1 \times 10^{-5}$
Longitude of the ascending node, $\Omega$ (°)	$326.459 \pm 0.001$	$311.81880 \pm 5 \times 10^{-5}$
Argument of perihelion, $\omega$ (°)	$107.67 \pm 0.17$	$120.55894 \pm 5 \times 10^{-5}$
Time of perihelion passage	2012 December 31.39 $\pm$ 0.17	2014 January 17.45565 $\pm$ $2 \times 10^{-5}$

The geocentric radiant and velocity of Chelyabinsk were  $\alpha_g = 22^\circ 11' 17'' \pm 26''$  ( $332.82 \pm 0.11^\circ$ ),  $\delta_g = +0^\circ 17' \pm 8''$  ( $+0.28^\circ \pm 0.14^\circ$ ),  $v_g = 15.14 \pm 0.16 \text{ km s}^{-1}$ . The orbit was obtained by numerical integration back in time to 60 d pre-impact, with the RADAU integrator<sup>28,29</sup>. The orbit of 86039 was taken from NASA JPL Database (<http://ssd.jpl.nasa.gov/sbdb.cgi>) on 7 June 2013. All angular coordinates are in equinox J2000.0.

acting on the body was computed at each fragmentation point to evaluate the mechanical strength of the body.

The first significant mass loss occurred at an altitude of about 45 km, under a dynamic pressure of 0.7 MPa. The series of most severe fragmentations occurred at altitudes between 40 and 30 km (1–5 MPa). Acoustic analysis revealed 11 individual fragmentations between altitudes 39.2 and 29.8 km (Extended Data Fig. 2), the two strongest at 31.9 and 30.6 km, respectively (with uncertainties of  $\pm 0.3$  km). By 29 km, the asteroid had already fragmented into about 20 large boulders of masses  $\sim 10^4$  kg. Judging from the deceleration, the mass of the leading boulder (main body) was  $\sim 2 \times 10^4$  kg. The boulders started to break up again at altitudes of 26–24 km (10–13 MPa). The main body remained intact and quite massive ( $\sim 10^4$  kg) until an altitude of 22 km, before severe disintegration began at a pressure of 18 MPa. Only a 15-kg fragment remained from the leading main mass at 17 km. Fragment F1, the largest individual piece surviving the descent, separated from one of the more decelerated boulders (not the main body) at an altitude of about 25 km, surviving a maximal dynamic pressure of 15 MPa at 20 km. Its trajectory deviated from the original direction of flight (Extended Data Fig. 3) by  $1.3^\circ \pm 0.4^\circ$ , implying that the lateral velocity gained at the break-up was  $400 \pm 130 \text{ m s}^{-1}$ . This lateral velocity is almost an order of magnitude larger than aerodynamic theory<sup>21</sup> and laboratory experiments<sup>22</sup> predict. Nevertheless, it is similar to the behaviour of the Morávka bolide<sup>10</sup>, suggesting that forces other than purely aerodynamic effects are also present during the fragmentation of bolides. Judging from its deceleration, the terminal mass of fragment F1 was  $450 \pm 50$  kg and dark-flight computation predicts a landing point in Lake Chebarkul, within 300 m of the actual impact site (Extended Data Fig. 4). The predicted

impact points and masses of other observed fragments (Extended Data Fig. 5) are given in Fig. 1 and Extended Data Fig. 6. Judging from the light-curve shape, the total mass of surviving fragments  $> 100$  g was at most only several per cent of the original mass and probably much less. It is much less than the pancake model predicts ( $> 50\%$ ) and less than the separated-fragments model predicts<sup>7</sup> ( $> 10\%$ ).

The fragmentation history shows that the bulk strength of the Chelyabinsk asteroid was  $\sim 1$  MPa, a value typical for smaller meteoroids, confirming that for near-Earth objects there is little correlation between strength and size on scales from centimetres to tens of metres<sup>6</sup>. Only the very strongest parts, representing  $< 1\%$  of the original mass, had strength  $> 15$  MPa, which may be comparable to the exceptionally strong Carancas meteoroid<sup>23,24</sup> and typical tensile strengths of recovered stony meteorites<sup>6</sup>. Even in the Martian atmosphere, the body would probably separate into several large fragments before reaching the surface. It is therefore not surprising that most fresh craters on Mars are found in clusters<sup>25</sup>.

A large dust trail was left in the atmosphere after the bolide passage. The southern and bottom sides of the trail were illuminated by the rising Sun. From northern sites the trail looked thin because only the illuminated bottom was visible, except for the first seconds, when the trail was self-luminous. The trail started at an altitude of 68 km. Between 60 and 26 km, the trail was thick, with radius 1.0–1.8 km. A thinner trail continued to 21.5 km and a very thin part extended to 18 km. The total volume of the trail was of order  $600 \text{ km}^3$  and its air mass was  $\sim 5 \times 10^9$  kg. The lower edge of the thick trail was almost stationary immediately after the passage of the bolide but then moved forwards in two waves as the material originally released at higher heights arrived. The velocities of these shocks were 2.8 and  $1.7 \text{ km s}^{-1}$ , respectively



**Figure 1 | Ground projection of the terminal part of the bolide trajectory and meteorite-strewn field.** Main trajectory (thick red line) and trajectory of fragment F1 (thin orange line) as plotted on Google Earth. The marks denote altitudes in kilometres. The predicted impact positions of 11 observed fragments (F1–F4, F6, F7 and F11–F15) are shown. The impact positions of simulated small (non-observable) fragments are also given. Yellow dots denote fragments that separated at lower altitude (21–26 km), pink fragments originate in the main break-ups at 30–39 km and brown fragments are from break-ups above 40 km. The three dot sizes correspond to terminal masses of 1–10 g, 10–100 g and  $> 100$  g. The second largest fragment, F2, had an estimated mass of 30 kg based on its observed deceleration. The dynamics after

luminous flight ceased<sup>30</sup> (dark flight) was computed using the wind field from the nearest available radiosonde at Verkhnee Dubrovo (180 km north of Chelyabinsk) measured at 0:00 UT on 15 February 2013. Using the wind field measured by radiosonde at Kurgan (250 km east of Chelyabinsk) would shift the meteorites much more to the south (by 2.5 km for a 200-g meteorite). The position of the impact hole in Lake Chebarkul (marked ‘Crater’) and the centroid of strewn field of small meteorites<sup>13</sup> are also shown. See also Extended Data Fig. 4. We expect that, as in other cases<sup>11</sup>, the mass distribution will be more complicated and the strewn field will be larger than our idealized model predicts.

(Extended Data Fig. 6). Larger regions of hot material within the trail continued moving forwards for some time after the bolide passage. The forward velocity of the brightest (and lowest) of several distinct hotspots (Extended Data Fig. 7) decreased from 0.7 to 0 km s<sup>-1</sup> during the period 1.5–4 s after the passage of the body. At the same time, a constant vertical velocity of 0.08 km s<sup>-1</sup> was measured for this hotspot (Extended Data Fig. 6). The upward motion, caused by buoyancy of the hot mixture of air, vapour and dust, continued when the forward motion stopped. About 15 s after the bolide passage, splitting of the trail into two parallel tracks became apparent (Extended Data Fig. 8), an effect seen in high-altitude luminous meteor trails and caused by convective instability leading to vortices<sup>26</sup>. After 50 s, the top of the cloud, which formed from the hotspot, was 3 km above the bottom of the trail (the trail as a whole moved about 0.5 km upwards). The maximum height of 6–7 km was reached after 3 min (Extended Data Fig. 9).

On the basis of the fragmentation strength of the Chelyabinsk asteroid, we conclude that it was probably a fractured single stone and not a rubble-pile assemblage, whose expected strength would be only ~25 Pa (ref. 27). The unusually close orbital association with asteroid 86039 suggests a possible genetic connection. Detailed comparison of reflectance spectra from the Chelyabinsk meteorite and 86039 are highly desirable to help explore this relationship. If the orbital association is real, we would also predict a very short exposure age for Chelyabinsk relative to other LL chondrites, because such orbital associations can only persist for a small fraction of the dynamical lifetime of a near-Earth asteroid.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 27 June; accepted 17 September 2013.**

**Published online 6 November 2013.**

- Vasilyev, N. V. The Tunguska meteorite problem today. *Planet. Space Sci.* **46**, 129–150 (1998).
- Boslough, M. B. E. & Crawford, D. A. Low altitude airbursts and the impact threat. *Int. J. Impact Eng.* **35**, 1441–1448 (2008).
- Silber, E. A., Le Pichon, A. & Brown, P. G. Infrasonic detection of a near-Earth object impact over Indonesia on 8 October 2009. *Geophys. Res. Lett.* **38**, L12201 (2011).
- McCord, T. B. *et al.* Detection of a meteoroid entry into the Earth's atmosphere on February 1, 1994. *J. Geophys. Res.* **100** (E2), 3245–3249 (1995).
- Brown, P. G. *et al.* A 500-kiloton airburst over Chelyabinsk and an enhanced hazard from small impactors. *Nature* <http://dx.doi.org/10.1038/nature12741> (this issue).
- Popova, O. *et al.* Very low strengths of interplanetary meteoroids and small asteroids. *Meteorit. Planet. Sci.* **46**, 1525–1550 (2011).
- Bland, P. A. & Artemieva, N. A. Efficient disruption of small asteroids by Earth's atmosphere. *Nature* **424**, 288–291 (2003).
- Jenniskens, P. *et al.* Radar-enabled recovery of the Sutter's Mill meteorite, a carbonaceous chondrites regolith breccia. *Science* **338**, 1583–1587 (2012).
- Borovička, J., Popova, O. P., Nemtchinov, I. V., Spurný, P. & Ceplecha, Z. Bolides produced by impacts of large meteoroids into the Earth's atmosphere: comparison of theory with observations. I. Benešov bolide dynamics and fragmentation. *Astron. Astrophys.* **334**, 713–728 (1998).
- Borovička, J. & Kalenda, P. The Morávka meteorite fall: 4. Meteoroid dynamics and fragmentation in the atmosphere. *Meteorit. Planet. Sci.* **38**, 1023–1043 (2003).
- Borovička, J. *et al.* The Košice meteorite fall: atmospheric trajectory, fragmentation, and orbit. *Meteorit. Planet. Sci.* <http://dx.doi.org/10.1111/maps.12078> (17 April 2013).
- Spurný, P. *et al.* The Bunburra Rockhole meteorite fall in SW Australia: fireball trajectory, luminosity, dynamics, orbit, and impact position from photographic and photoelectric records. *Meteorit. Planet. Sci.* **47**, 163–185 (2012).
- Nazarov, M. A. Chelyabinsk. *Meteorit. Bull.* **102** (2013).
- Borovička, J. The comparison of two methods of determining meteor trajectories from photographs. *Bull. Astron. Inst. Czechoslovakia* **41**, 391–396 (1990).
- Delbó, M., Harris, A. W., Binzel, R. P., Pravec, P. & Davies, J. K. Keck observations of near-Earth asteroids in the thermal infrared. *Icarus* **166**, 116–130 (2003).
- Binzel, R. P. *et al.* Observed spectral properties of near-Earth objects: results for population distribution, source regions, and space weathering processes. *Icarus* **170**, 259–294 (2004).
- Southworth, R. B. & Hawkins, G. S. Statistics of meteor streams. *Smithsonian Contrib. Astrophys.* **7**, 261–285 (1963).
- Drummond, J. D. A test of comet and meteor shower associations. *Icarus* **45**, 545–553 (1981).
- Mainzer, A. *et al.* NEOWISE observations of near-Earth objects: preliminary results. *Astrophys. J.* **743**, 156 (2011).
- Botke, W. F. Jr *et al.* Debaised orbital and absolute magnitude distribution of the near-Earth objects. *Icarus* **156**, 399–433 (2002).
- Artemieva, N. A. & Shuvalov, V. V. Motion of fragmented meteoroid through the planetary atmosphere. *J. Geophys. Res.* **106** (E2), 3297–3309 (2001).
- Park, C. & Brown, J. D. Fragmentation and spreading of a meteor-like object. *Astron. J.* **144**, 184 (2012).
- Brown, P. *et al.* Analysis of a crater-forming meteorite impact in Peru. *J. Geophys. Res.* **113**, E09007 (2008).
- Borovička, J. & Spurný, P. The Carancas meteorite impact: encounter with a monolithic meteoroid. *Astron. Astrophys.* **485**, L1–L4 (2008).
- Daubar, I. J., McEwen, A. S., Byrne, S., Kennedy, M. R. & Ivanov, B. The current Martian cratering rate. *Icarus* **225**, 506–516 (2013).
- Kelley, M. C., Williamson, C. H. K. & Vlasov, M. N. Double laminar and turbulent meteor trails observed in space and simulated in the laboratory. *J. Geophys. Res.* **118**, 3622–3625 (2013).
- Sánchez, P. & Scheeres, D. J. The strength of regolith and rubble pile asteroids. Preprint at <http://arxiv.org/abs/1306.1622> (2013).
- Everhart, E. in *Dynamics of Comets: Their Origin and Evolution* (eds Carusi, A. & Valsecchi, G.) 185 (Kluwer, 1985).
- Clark, D. & Wiegert, P. A numerical comparison with the Ceplecha analytical meteoroid orbit determination method. *Meteorit. Planet. Sci.* **46**, 1217–1225 (2011).
- Ceplecha, Z. Geometric, dynamic, orbital and photometric data on meteoroids from photographic fireball networks. *Bull. Astron. Inst. Czechoslovakia* **38**, 222–234 (1987).

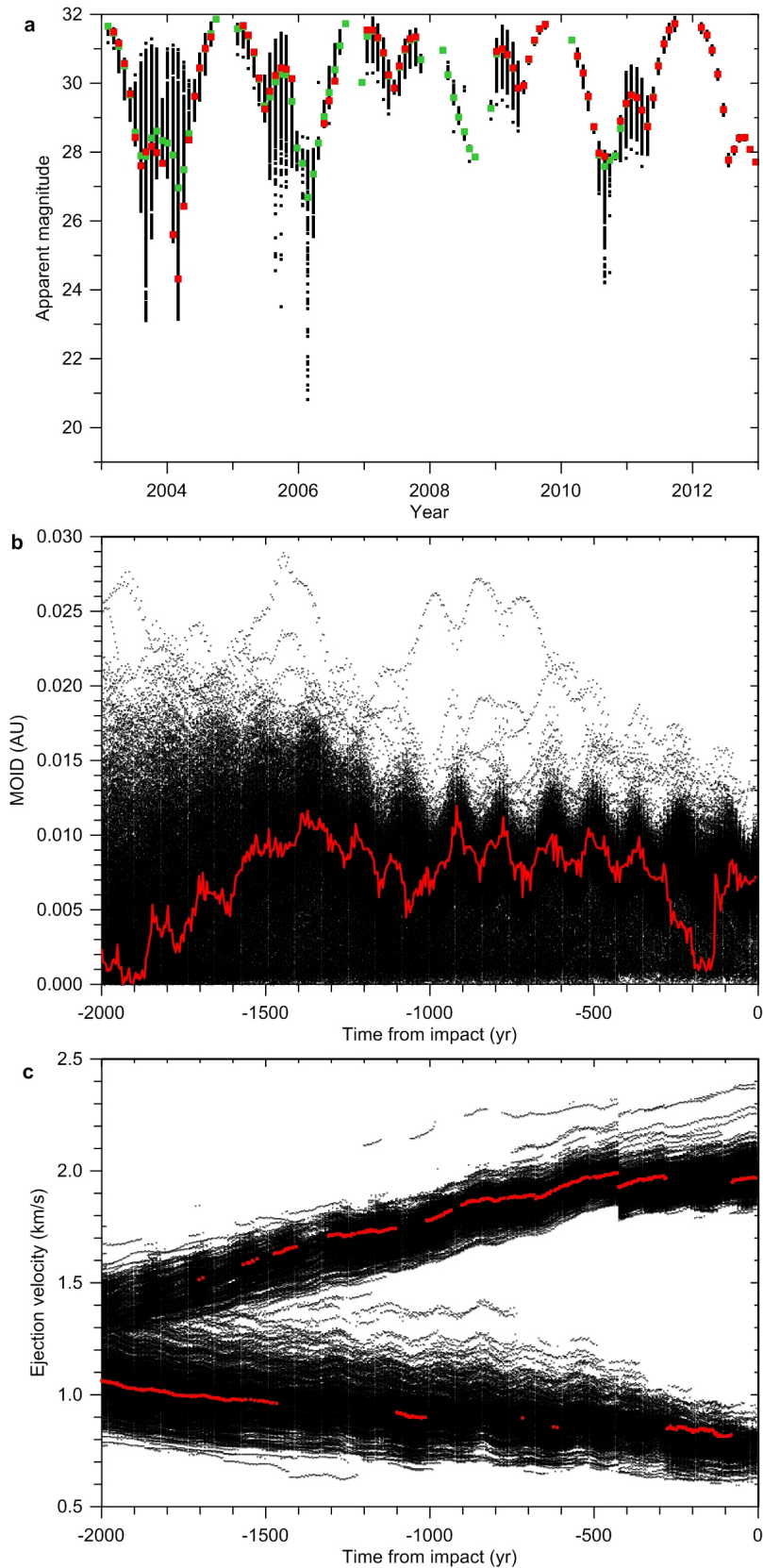
**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank D. Částeček and O. Popova and her team (V. Emelyanenko, A. Kartashova, D. Glazachev and E. Biryukov) for providing the nocturnal *in situ* calibration images. We are obliged to all the videographers who posted videos of the Chelyabinsk superbolide on the internet. The work of J.B., P.S. and L.S. was supported by grant no. P209/11/1382 from GACR and Praemium Academiae. The Czech institutional project was RVO:67985815. The work of P.B., P.W. and D.C. was supported in part by the Natural Sciences and Engineering Research Council of Canada and NASA's Meteoroid Environment Office.

**Author Contributions** J.B. made measurements from most of the videos, computed the bolide trajectory and velocity, and analysed its atmospheric fragmentation and dust trail. P.S. organized the calibrations, made measurements from the calibration images and participated in interpreting the results. P.B. participated in the acoustic analysis and in interpreting the results. P.W. and D.C. performed the orbital integration, analysed the parent-body linkage and analysed the asteroid visibility before impact. P.K. found many important videos and participated in the acoustic analysis. L.S. prepared the calibrations and participated in video measurements. All authors commented on the manuscript.

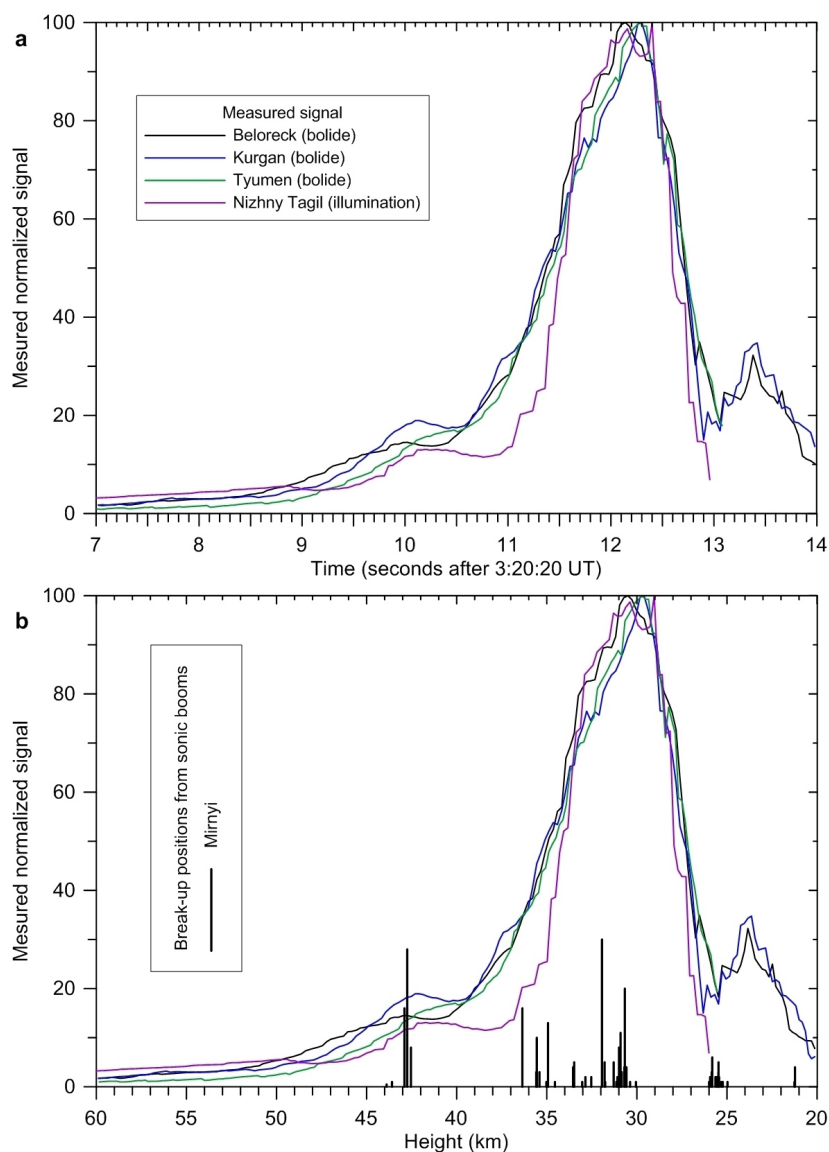
**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.B. ([jiri.borovicka@asu.cas.cz](mailto:jiri.borovicka@asu.cas.cz)).





**Extended Data Figure 1 | Visibility and orbital evolution of Chelyabinsk asteroid in the past.** The results of backward integration of Chelyabinsk nominal orbit (red) and its 1,000 clones (black dots). **a**, Apparent magnitude as seen from the Earth at 30-day intervals during past 10 years. Green, mean of all clones. Plotted only for elongations  $>45^\circ$  from the Sun. **b**, Minimum orbit

intersection distance (MOID) between the Chelyabinsk orbit and the osculating orbit of asteroid 86039 during the past 2,000 years. **c**, Change in velocity required to reach Chelyabinsk orbit from the orbit of 86039 at MOID during the past 2,000 years.



c

**Heights of break-ups from sonic booms at 19 videos**

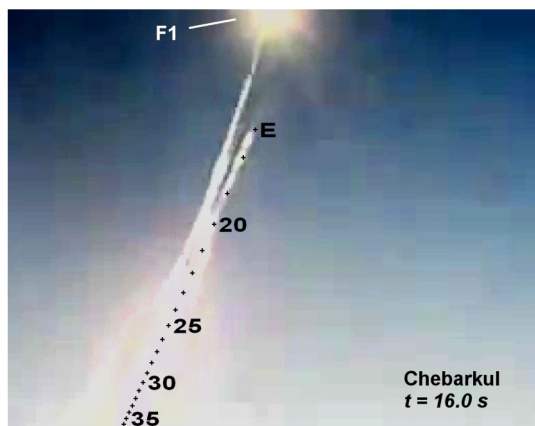
42.7 <sup>†</sup> - 39.2 <sup>†</sup> - 37.2 <sup>†</sup> - 36.4 - 35.5 - 34.9 - 33.5 - 32.6 - <b>31.9*</b> - 30.9 - <b>30.6*</b> - 29.8
25.8 - 25.5 - 21.3 - 18.6 <sup>†</sup> - 17.5 <sup>†</sup>

\*strong

<sup>†</sup>not confirmed on video A19<sup>‡</sup>only on video A19 but confirmed on seismic station BRVK. Another break-up suggested by BRVK at ~ 47 km.

**Extended Data Figure 2 | Light curve of Chelyabinsk superbolide in relative units and fragmentation altitudes as determined from sonic booms.** The luminous signal was computed in relative units from pixel sum values from substantial parts of the images, and then normalized to 100. Corrections to bolide range and atmospheric extinction were applied but no attempt to convert the signal to absolute units was made (for the absolute light curve, see ref. 5). For each video, the measured pixel sum was corrected using the estimated changes of automatic gain control of the camera. The absolute timing was obtained from the Nizhny Tagil video (L1) and the height scale from the Beloreck video (video 14, or L4). The fragmentation altitudes were

determined from the timing of secondary sonic booms and numerical ray-tracing modelling of the sonic wave's propagation from the bolide to the video sites. The videos used are listed in Extended Data Table 1. **a**, Bolide light curve as a function of time. **b**, The same data as a function of height compared with the computed source heights of sonic booms detected (as image failures) in the Mirnyi video (A19). The fragmentations are marked by vertical bars at the corresponding height. The length of the bar is proportional to the number of video frames affected by the failure. **c**, The compilation of sonic boom source heights from all 19 videos used for acoustic analysis.



**Extended Data Figure 3 | Deviation of fragment F1 from the main trajectory.** Frame from video 15. The time is counted from 3:20:20 UT. The labelled marks identify points on the main trajectory at the given altitude (in kilometres). E represents the endpoint of the main trajectory.



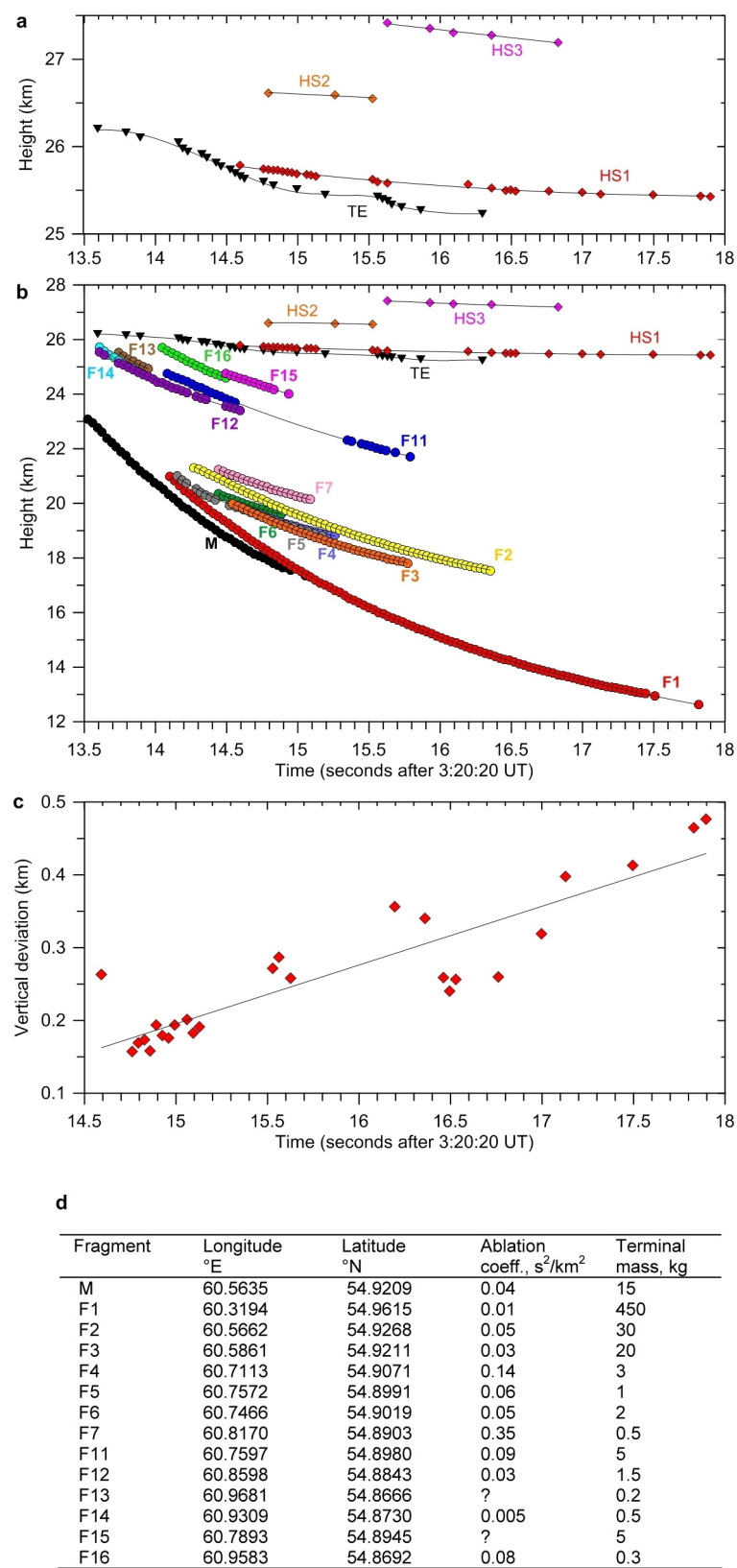


**Extended Data Figure 4 | Predicted impact position of fragment F1, computed with four different wind fields, compared with the position of the hole in the ice ('Crater').** The point marked F1 was computed with Verkhnee Dubrovo radiosonde data (0:00 UT). Point K is for Kurgan radiosonde (0:00 UT), point U is for the UKMO wind model for Chelyabinsk

(12:00 UT) and point G is for the G2S model (3:00 UT) (ref. 31 in Supplementary Information). The distance between U and K is 960 m. The distance between F1 and the crater is 220 m. We note that the position of the crater was not used for the computation of the F1 trajectory and impact point. The background image is from Google Earth and was taken one day after the impact.



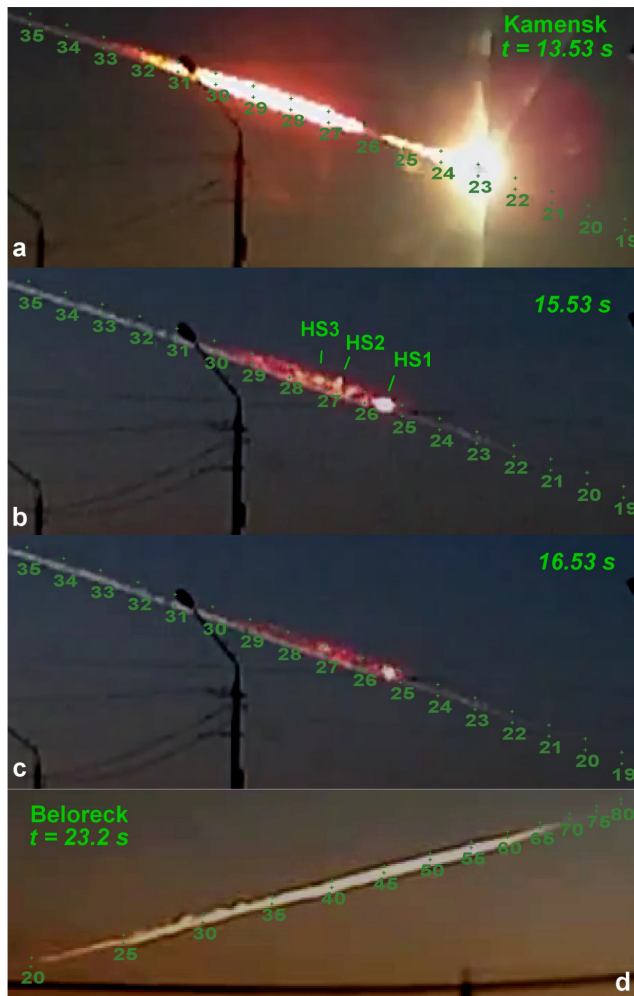
**Extended Data Figure 5 | Identification of fragments in a series of images from video 7.** Fragments F1–F7 originated at lower altitudes ( $\sim 25$  km), whereas fragments F11–F16 originated at higher altitudes ( $> 30$  km).



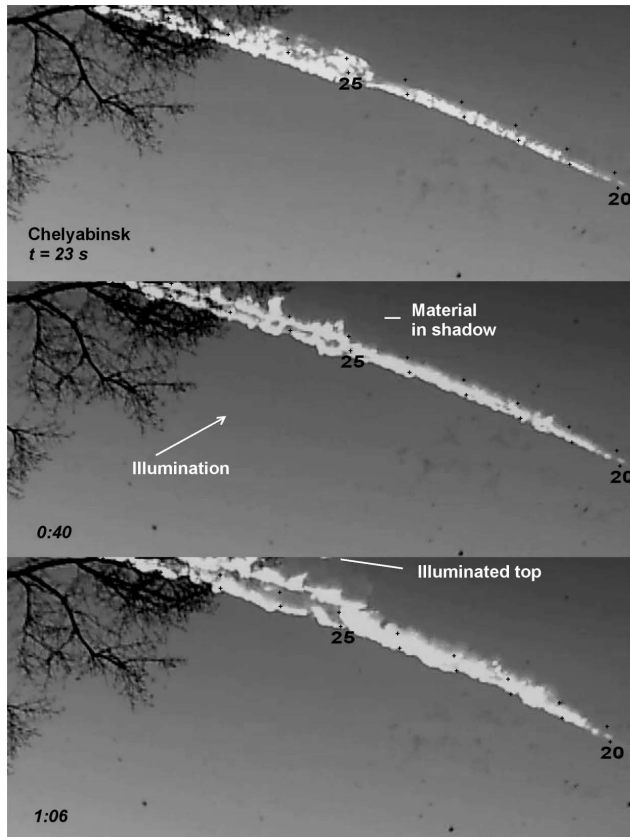
**Extended Data Figure 6 | Dynamics of the dust trail and fragments and predicted impact positions of observed fragments.** **a**, Altitude as a function of time for the lower edge of the thick dust trail (TE) and hotspots within the trail (HS1–HS3). The hotspots are identified in Extended Data Fig. 7. **b**, Altitude as a function of time for the main body (M), lower fragments (F1–F7) and upper fragments (F11–F16), plotted together with the dust trail features. The fragments are identified in Extended Data Fig. 5. The main body and trail were measured primarily from video 2; and the fragments, from video 7.

**c**, Upward motion of the main hotspot (HS1) within the dust trail. Vertical deviation of the centre of the hotspot from the trajectory is plotted against time. The linear fit gives upward velocity of  $0.08 \text{ km s}^{-1}$ . **d**, Predicted impact positions and dynamic properties of observed fragments. Ablation coefficients and terminal masses were obtained by fitting the observed decelerations. Masses are valid for assumed spherical shapes and bulk densities of  $3,300 \text{ kg m}^{-3}$ . In some cases, the ablation coefficient could not be computed because there was an insufficient number of data points.





**Extended Data Figure 7 | Images of the dust trail at early stages.** a–c, Images from a single video site (video 2) located north of the fireball trajectory. Time is counted from 3:20:20 UT. Three distinct hotspots (HS1–HS3) are identified. The labelled marks identify points on the trajectory at the given altitude (in kilometres). The unlabelled marks above them identify points at the same geographic coordinates but 1 km higher. They are provided to assess the width of the trail. d, Image from video 14, from the southwest. It demonstrates that the width of the fully illuminated fresh trail was  $\sim 2$  km over much of its length. For later evolution of the trail, see Extended Data Figs 8 and 9.



**Extended Data Figure 8 | Evolution of the lower part of the dust trail as seen from Chelyabinsk during the first minute.** Three frames from video 6 (the video has colour defects). The time is given in minutes and seconds and is counted from 3:20:20 UT. The lower marks identify points on the trajectory in 1-km altitude intervals. The upper marks identify points at the same geographic coordinates but 1 km higher. The video demonstrates vertical ascent and splitting of the trail. When the original video is speeded up, rotation of the material in the trail is clearly visible. The trail was illuminated from below. The ‘bubble’ formed at the position of the main hotspot (HS1; see Extended Data Fig. 7) was in shadow most of the time. Only its illuminated top is visible on the third frame, just at the edge of the field of view.



**Extended Data Figure 9 | Longer-term evolution of the dust trail.**

Five frames from an uncalibrated video (<http://www.youtube.com/watch?v=Z20lnOVscpc>, author D. Beletsky) taken from south of the fireball trajectory (on the road from Magnitogorsk to Chelyabinsk). The time is given in minutes and seconds and is counted from 3:20:20 UT. The trail was fully illuminated from this site. The video demonstrates the rise of the 'bubble' formed at the position of the main hotspot (HS1; see Extended Data Fig. 7). The maximum altitude was reached about 3 min after the bolide had passed by.



Extended Data Table 1 | List of YouTube videos used

Extended Data Table 1 | List of used YouTube videos

No.	City	Longitude °E	Latitude °N	Altitude m	YouTube code*	Author
<b>Videos used for trajectory determination</b>						
1	Verkhnyaya Pyshma	60.6082	56.9635	276	LzvipPOpUy0	Vos'moy Rayon
2	Kamensk Ural'sky	61.9186	56.4151	170	kFlpCT3v12E	Aleksandr Ivanov
3	Kamensk Ural'sky	61.9333	56.3933	151	7TPDwSXaiB0	LANCER96RUS
4	Kamensk Ural'sky	61.9426	56.3852	163	J3DqsbxKOMA	LANCER96RUS
5	Chelyabinsk	61.2967	55.2204	246	4ZxXYscmgRg	Andrey Borisovich Korolev
6	Chelyabinsk	61.4720	55.1797	231	OM-5ngYg5Mg	Vyacheslav Kravchenko
7	Chelyabinsk	61.3935	55.1756	228	8Eu7QAP2DPM	Viktor Borzov
8	Chelyabinsk	61.4448	55.1663	231	32aJ4RB8Mql	MegaProfitroll
9	Chelyabinsk	61.3637	55.1500	257	gQ6Pa5Pv_io	Dmitriy Volkov
10	Korkino	61.3995	54.8909	241	odKjwrjIM-k	nek rozato
11	Yemanzhelinsk	61.3040	54.7566	234	2Gc1HgO5hNY	Aleksandr Zakharov
12	Kichigino	61.2717	54.5008	238	0CoP7WB8Gew	Mikhail Troitsk
13	Troitsk	61.5313	54.0771	183	UjNpJXP7trQ	C2Crash
14	Beloreck	58.4102	53.9527	532	5_1ytDqps8A	MrKuzaman
15	Chebarkul	60.4002	54.9950	343	xboo7LiNR08	yulya Ryzhaya
<b>Videos used for light curve measurement</b>						
L1	Nizhny Tagil	59.9439	57.8703	220	NcZNhJVW5xl†	SetiTagila
L2	Tyumen'	65.6053	57.2024	60	1ZdYf2vM5LA	EastSide287
L3	Kurgan	65.2956	55.4735	70	X2ja6_zJtzk	Dmitriy Grekov
L4	Beloreck	58.4102	53.9527	532	5_1ytDqps8A	MrKuzaman
<b>Videos used for acoustic analysis</b>						
A1	Chelyabinsk	55.1647	61.5482		HaMurqKMenw	Alexandr Gubarev
A2	Malinovka	55.0900	61.2500		5Rh2-v-gFEs	Svladislav74
A3	Pervomaysky	54.8726	61.2008		R99zvcrqXo8	Axel Alex
A4	Chelyabinsk	55.2675	61.4082		hMZkv0-2500	NIKI4174
A5	Chelyabinsk	55.1799	61.3479		5xMOSY4bW_M	Aleksej D
A6	Chelyabinsk	55.1570	61.3657		uXU3z3-bxNk	Andrey Yurkin
A7	Chelyabinsk	55.1500	61.3637		gQ6Pa5Pv_io	Dmitriy Volkov
A8	Chelyabinsk	55.1114	61.3509		Np_mpGYSBsA	Serg Kh
A9	Chelyabinsk	55.1650	61.4070		rftTN4XAt34	Maxim Savelyev
A10	Chelyabinsk	55.1582	61.4113		G2KpK_GmvA8	Yuriy Bazhaev
A11	Chelyabinsk	55.1570	61.4452		38zQkZCWIL0	Sergey Polyakov
A12	Chelyabinsk	55.1636	61.4703		zkVUbMdAV2Q	Alexandr Burlakov
A13	Chelyabinsk	55.1665	61.4639		QTZ0vesq5fA	greshnikacw
A14	Yemanzhelinsk	54.7564	61.3280		KmjJGgvOIMY	Daniil Lysenko
A15	Chelyabinsk	55.1404	61.4780		16KxOxjSndA	dima stepanov
A16	Kopeysk	55.1210	61.6046		VZJEfNld_JU	Aleksandr Al'shevsky
A17	Kopeysk	55.1103	61.6042		4x3xG-eKpNA	ZaTDro
A18	Kopeysk	55.1087	61.6183		UsOEy5hCCow	Andrey Kostomarov
A19	Mirnyi	54.577	60.312		LmlwWFDylx0	kaban0796

\* <http://www.youtube.com/watch?v=code>

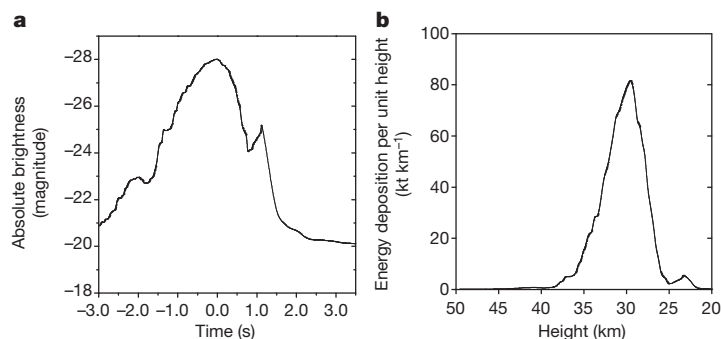
† The last of the four videos in the compilation was used

# A 500-kiloton airburst over Chelyabinsk and an enhanced hazard from small impactors

P. G. Brown<sup>1,2</sup>, J. D. Assink<sup>3</sup>, L. Astiz<sup>4</sup>, R. Blaauw<sup>5</sup>, M. B. Boslough<sup>6</sup>, J. Borovička<sup>7</sup>, N. Brachet<sup>3</sup>, D. Brown<sup>8</sup>, M. Campbell-Brown<sup>1</sup>, L. Ceranna<sup>9</sup>, W. Cooke<sup>10</sup>, C. de Groot-Hedlin<sup>4</sup>, D. P. Drob<sup>11</sup>, W. Edwards<sup>12</sup>, L. G. Evers<sup>13,14</sup>, M. Garces<sup>15</sup>, J. Gill<sup>1</sup>, M. Hedlin<sup>4</sup>, A. Kingery<sup>16</sup>, G. Laske<sup>4</sup>, A. Le Pichon<sup>3</sup>, P. Mialle<sup>8</sup>, D. E. Moser<sup>5</sup>, A. Saffer<sup>10</sup>, E. Silber<sup>1</sup>, P. Smets<sup>13,14</sup>, R. E. Spalding<sup>6</sup>, P. Spurný<sup>7</sup>, E. Tagliaferri<sup>17</sup>, D. Uren<sup>1</sup>, R. J. Weryk<sup>1</sup>, R. Whitaker<sup>18</sup> & Z. Krzeminski<sup>1</sup>

Most large (over a kilometre in diameter) near-Earth asteroids are now known, but recognition that airbursts (or fireballs resulting from nuclear-weapon-sized detonations of meteoroids in the atmosphere) have the potential to do greater damage<sup>1</sup> than previously thought has shifted an increasing portion of the residual impact risk (the risk of impact from an unknown object) to smaller objects<sup>2</sup>. Above the threshold size of impactor at which the atmosphere absorbs sufficient energy to prevent a ground impact, most of the damage is thought to be caused by the airburst shock wave<sup>3</sup>, but owing to lack of observations this is uncertain<sup>4,5</sup>. Here we report an analysis of the damage from the airburst of an asteroid about 19 metres (17 to 20 metres) in diameter southeast of Chelyabinsk, Russia, on 15 February 2013, estimated to have an energy equivalent of approximately 500 (±100) kilotons of trinitrotoluene (TNT, where 1 kiloton of TNT =  $4.185 \times 10^{12}$  joules). We show that a widely

referenced technique<sup>4–6</sup> of estimating airburst damage does not reproduce the observations, and that the mathematical relations<sup>7</sup> based on the effects of nuclear weapons—almost always used with this technique—overestimate blast damage. This suggests that earlier damage estimates<sup>5,6</sup> near the threshold impactor size are too high. We performed a global survey of airbursts of a kiloton or more (including Chelyabinsk), and find that the number of impactors with diameters of tens of metres may be an order of magnitude higher than estimates based on other techniques<sup>8,9</sup>. This suggests a non-equilibrium (if the population were in a long-term collisional steady state the size-frequency distribution would either follow a single power law or there must be a size-dependent bias in other surveys) in the near-Earth asteroid population for objects 10 to 50 metres in diameter, and shifts more of the residual impact risk to these sizes.



**Figure 1 | Light curve of the Chelyabinsk airburst.** **a**, The brightness profile for the Chelyabinsk airburst, based on indirect illumination measured from video records. The brightness is an average derived from indirect scattered sky brightness from six videos proximal to the airburst, corrected for the sensor gamma setting, autogain, range and airmass extinction, following the procedure used for other airburst light curves generated from video<sup>24,25</sup>. The light curve has been normalized using the US government sensor data peak brightness value of  $2.7 \times 10^{13} \text{ W sr}^{-1}$ , corresponding to an absolute astronomical magnitude of  $-28$  in the silicon bandpass. The individual video light curves deviate by less than one magnitude between times  $-2$  and  $+1.5$  with larger deviations outside this interval. Time zero corresponds to 03:20:32.2 UTC on 15 February 2013. **b**, The energy deposition per unit height

for the Chelyabinsk airburst, based on video data. The conversion to absolute energy deposition per unit path length assumes a blackbody emission of 6,000 K and bolometric efficiency of 17%, the same as the assumptions used to convert earlier US government sensor information to energy<sup>26</sup>. The heights are computed using the calibrated trajectory<sup>10</sup> and features of the light curves common to different video sites, resulting in a height accuracy of about 1 km. The total energy of the airburst found by integrating under the curve exceeds 470 kt. The half-energy-deposition height range is 33–27 km; these are the heights at which energy deposition falls below half the peak value of approximately 80 kt per kilometre of height, which is reached at an altitude near 29.5 km.

<sup>1</sup>Department of Physics and Astronomy, University of Western Ontario, London, Ontario N6A 3K7, Canada. <sup>2</sup>Centre for Planetary Science and Exploration, University of Western Ontario, London, Ontario N6A 5B7, Canada. <sup>3</sup>Commissariat à l'Energie Atomique, Département Analyse Surveillance Environnement (CEA/DAM/DIF), Bruyères-le-Châtel, 91297 Arpajon Cedex, France. <sup>4</sup>Laboratory for Atmospheric Acoustics, Institute of Geophysics and Planetary Physics, University of California, San Diego, La Jolla, California 92093-0225, USA. <sup>5</sup>Marshall Information Technology Services (MITS)/Dynamics Technical Services, NASA Marshall Space Flight Center, Huntsville, Alabama 35812, USA. <sup>6</sup>Sandia National Laboratories, PO Box 5800, Albuquerque, New Mexico 87185, USA. <sup>7</sup>Astronomical Institute, Academy of Sciences of the Czech Republic, CZ 251 65 Ondřejov, Czech Republic. <sup>8</sup>International Data Center, Provisional Technical Secretariat, Comprehensive Test Ban Treaty Organization, PO Box 1200, A-1400 Vienna, Austria. <sup>9</sup>Bundesanstalt für Geowissenschaften und Rohstoffe, Stilleweg 2, 30655 Hannover, Germany. <sup>10</sup>Meteoroid Environments Office, EV44, Space Environment Team, Marshall Space Flight Center, Huntsville, Alabama 35812, USA. <sup>11</sup>Space Science Division, Naval Research Laboratory, 4555 Overlook Avenue, Washington DC 20375, USA. <sup>12</sup>Natural Resources Canada, Canadian Hazard Information Service, 7 Observatory Crescent, Ottawa, Ontario K1A 0Y3, Canada. <sup>13</sup>Seismology Division, Royal Netherlands Meteorological Institute, Wilhelminalaan 10, 3732 GK De Bilt, The Netherlands. <sup>14</sup>Department of Geoscience and Engineering, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands. <sup>15</sup>Infrasound Laboratory, University of Hawaii, Manoa 73-4460 Queen Kaahumanu Highway, 119 Kailua-Kona, Hawaii 96740-2638, USA. <sup>16</sup>ERC Incorporated/Jacobs ESSSA Group, NASA Marshall Space Flight Center, Huntsville, Alabama 35812, USA. <sup>17</sup>ET Space Systems, 5990 Worth Way, Camarillo, California 93012, USA. <sup>18</sup>Los Alamos National Laboratory, EES-17 MS F665, PO Box 1663 Los Alamos, New Mexico 87545, USA.

The Chelyabinsk airburst<sup>10</sup> was observed globally by multiple instruments—including infrasound, seismic, US government sensors and more than 400 video cameras—at ranges up to 700 km away. The resulting airblast (shock wave travelling through the air from an explosion) shattered thousands of windows in urban Chelyabinsk, with flying glass injuring many residents.

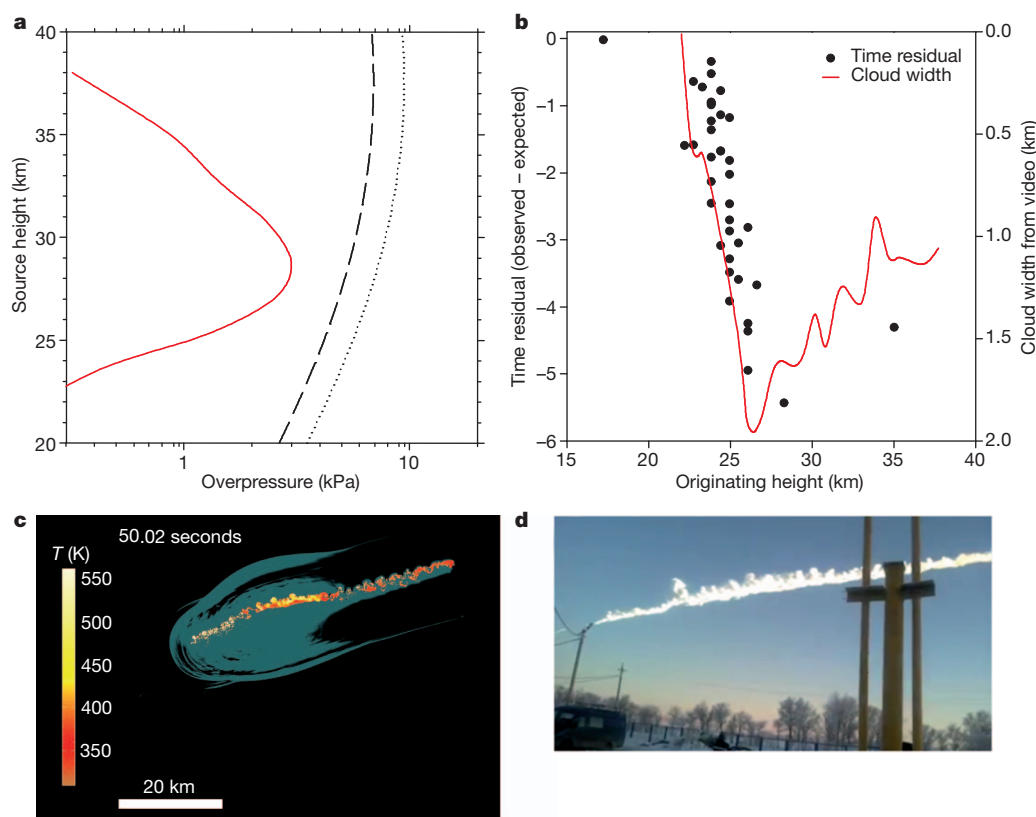
Data from US government sensors timed the peak brightness to 03:20:32.2 UTC (coordinated universal time) on 15 February 2013 with an integrated radiated energy of  $3.75 \times 10^{14}$  J and a peak brightness of  $2.7 \times 10^{13}$  W sr<sup>-1</sup>. These values correspond to an estimated energy equivalent of about 530 kt of TNT. The peak brightness was equivalent to an absolute astronomical magnitude of  $-28$  (referenced to a range of 100 km) in the silicon bandpass, making the airburst appear 30 times brighter than the Sun to an observer directly under this point. The airburst's light curve has been reconstructed by considering the measured light production from several video records (see Supplementary Information for details) as shown in Fig. 1. We note that point-like models<sup>4–6</sup> of airburst energy deposition, which treat the impactor as a strengthless, liquid-like material, predict that the height range in which the energy deposition per unit path length falls to half its maximum value is less than 2 km for impacts as shallow ( $17^\circ$  from the horizontal)<sup>10</sup>

**Table 1 | Energy estimates for the Chelyabinsk airburst**

Technique	Best estimate (kt)	Range (kt)
Seismic	430	220–630
Infrasound (mean period)	600	350–990
US government sensor	530	450–640
Video-derived lightcurve	>470	

Here 'kiloton' refers to the energy equivalent to a kiloton of TNT. To estimate the energy from infrasonic airwaves, all 42 infrasound stations of the International Monitoring System<sup>23</sup> were examined. Of these, 20 stations showed clear signals from the airburst. Our infrasound energy estimates are based on the average observed dominant infrasound period from 12 stations that have stratospheric returns showing the highest signal-to-noise ratio. Seismic Rayleigh waves generated by the airburst shock wave impinging on the Earth's surface just south of Chelyabinsk were detected by about 70 seismic stations at ranges in excess of 4,000 km. The amplitude of these waves in specific passbands as calibrated to nuclear airbursts<sup>19</sup> were used as an independent estimate of source energy.

as that of Chelyabinsk, which is less than one-third of the observed value (more than 6 km). (We note that any object striking the Earth or its atmosphere is an impactor; a ground impactor creates a crater, but most burn up before that, releasing a large amount of energy into the atmosphere as an airburst.) Airburst energy estimates from four different techniques are summarized in Table 1. Our preferred mean energy estimate is in the range of 400–600 kt. Details of the analysis



**Figure 2 | Observed and predicted shock characteristics for the Chelyabinsk airburst.** **a**, Theoretical airblast overpressures using standard nuclear weapons relations<sup>7</sup> and the cylindrical-line source blast theory<sup>13</sup> (which assumes the explosion occurs so swiftly that it is equivalent to a single instantaneous detonation of a long cylindrical line of explosive charge) that is appropriate to central Chelyabinsk. The nuclear relation curves (in black) assume a spherical point source at a specific height and show assumed yields of 500 kt (dashed line) and 1 Mt (dotted line). The cylindrical-line source airblast model (red line) uses the energy deposition per unit length from Fig. 1b to define an equivalent blast radius as the source and assumes that the shock is linear at the ground (linear means its amplitude is low enough to be well approximated as moving at the local ambient speed of sound and non-linear effects are negligible). **b**, Travel-time residuals between the time of airburst passage at each height and the main airblast arrival for 38 videos (see Supplementary Table 5). The residuals (black circles) show the observed arrival time (corrected for fireball

motion) minus the expected time, calculated assuming propagation at the local adiabatic sound speed and incorporating winds<sup>11</sup>. For comparison, the width of the visible cloud trail is shown (red line). This is consistent with the shock wave travelling faster than the ambient sound speed near the airburst. The minimum timing residuals suggest that shock source heights vary between 30 km and 23 km across Chelyabinsk, as opposed to originating from a point source. **c**, Modelling the temperature of the trail (colour scale). The apparent cross-section of CTH simulation 50 s after 0.5 Mt was released into segmented cylinders of air. The outer blue-grey shaded area reveals an envelope of shocked air—note that the dominant shape of the shock is cylindrical. **d**, For comparison with **c**, a video frame of the dust cloud taken 250 km to the southwest of the airburst path looking North is shown (<http://www.youtube.com/watch?v=ICv9S0Z0e0E>, taken by E. Volkov). Approximately 140 km of the end portion of the airburst trail is shown, some 40–60 s after the passage of the fireball.



procedures and measurements are given in the Supplementary Information.

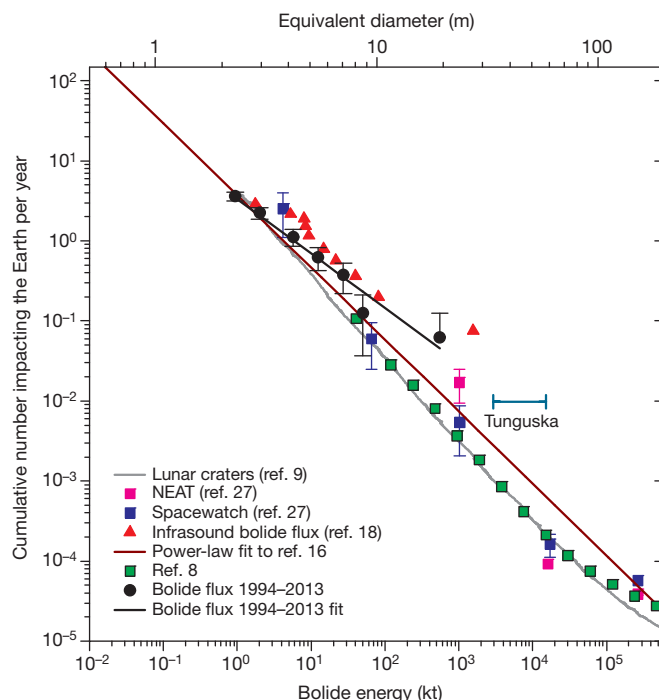
To establish the nature and source height of the airblast that caused damage in Chelyabinsk, we used the known trajectory<sup>10</sup> and a suite of videos (see Supplementary Table 5) that recorded both the airburst and the main airblast arrival. We computed acoustic travel times from each point on the airburst trajectory to each video location using a propagation model including winds that was developed for earlier airburst infrasound analyses<sup>11</sup>. The results show that the first airblast wave (which also produced the damage) arrived from different altitudes at different sites, consistent with a cylindrical shock from the extended airburst, as opposed to a more point-like explosion. The timing residuals between the observed and expected arrivals follow the bolide trail size, as shown in Fig. 2b, consistent with the shock being strong early in its propagation. The airblast reaching the city of Chelyabinsk was generated at altitudes of 24–30 km, roughly from the peak to the end of the main airburst.

In Fig. 2a we show overpressure predictions from standard airblast relations based on nuclear explosions<sup>7</sup>, as used by most impact-effect models<sup>4,6,12</sup>. For comparison, the predictions of cylindrical-line source airblast theory applied to meteor entry<sup>13</sup> are also shown. The airblast overpressure in Chelyabinsk from window breakage measurements is  $3.2 \pm 0.6$  kPa (see Supplementary Information for details). We note the overestimation of overpressure using the nuclear blast relations<sup>7</sup>, an effect others have suggested in connection with airbursts<sup>4</sup>. Given that nuclear explosions release half their energy as radiation<sup>7</sup>, thus reducing the effective yield of airblast energy, the nuclear curve in Fig. 2a that is most appropriate to Chelyabinsk is about 1 Mt.

To examine whether a fragmentation model<sup>14</sup> is consistent with the observed data and estimated object size, we have applied an entry code based on a progressive fragmentation model of the initial object. Assuming an initial meteoroid of diameter 19 m and a tensile strength at first fragmentation of 0.7 MPa (ref. 10), with ablation ending at about 27 km once most of the energy has been lost, we find a reasonable match to both the light curve and early dynamics. The final main fragmentations in this model occur near 4 MPa, very similar to those observed (1–5 MPa) in the most severe fragmentation portion of the airburst<sup>10</sup>. The dynamics and light production from the model are not realistic near the terminal phase of the airburst because the model assumes that all fragments split identically at each fragmentation epoch. This is in contrast to observations at the end of the airburst where one leading fragment was observed<sup>10</sup> (as opposed to dozens of identically sized individuals).

To further define the nature of the shock, we have used the well known CTH simulation framework used for the Tunguska<sup>15</sup> airburst and impactors<sup>1</sup> comparable to the asteroid causing the Chelyabinsk airburst. The simulation used all the observed trajectory parameters<sup>10</sup> and the observed energy as a function of height (Fig. 1b) to mimic the entry process by creating an instantaneous energy release in a sequence of momentum-preserving air cylinders along the airburst path, scaled such that the total integrated energy is 500 kt. Figure 2c and d shows the result of this simulation and comparison to a video record of the dust cloud generated by the airburst at a similar time. The notable characteristics are that the primary shock is cylindrical, in contrast to point-source energy release airburst models<sup>4–6</sup>, which have a strong spherical shock component. Instabilities that result from fast-rising buoyant air in the simulation produce similar structures to those seen in videos of the dust cloud.

Model overpressures for central Chelyabinsk are found to be 3 kPa, consistent with observations. Our estimates of overpressure are based on window breakage (see Supplementary Fig. 5) confined to a small region in Chelyabinsk. The CTH simulations were run for more than three minutes after the airburst, producing model variations of overpressure across the entire city of Chelyabinsk which were smaller than the differences produced by local effects, such as shock reflections from buildings, numerical uncertainty in the simulation and our generally



**Figure 3 | The estimated cumulative flux of impactors at the Earth.** The bolide impactor flux at the Earth (bolide flux 1994–2013; black circles) is based on about 20 years of global observations from US government sensors and infrasound airwave data. Global coverage averages 80% among a total of 58 observed bolides with  $E > 1$  kt and includes the Chelyabinsk bolide (rightmost black circle). This coverage correction is approximate and the bolide flux curve is probably a lower limit. The brown line represents an earlier power-law fit from a smaller data set for bolides 1–8 m in diameter<sup>16</sup>. Error bars represent counting statistics only. For comparison, we plot de-biased estimates of the near-Earth-asteroid impact frequency based on all asteroid survey telescopic search data until mid-2012 (green squares)<sup>8</sup> and other earlier independently analysed telescopic data sets<sup>27</sup> including the NEAT discoveries (pink squares) and the Spacewatch survey (blue squares), where diameters are determined assuming an albedo of 0.1. From the telescopically determined number of near-Earth asteroids and their typical orbits we can compute the average interval between Earth impactors of a given energy. Energy for telescopic data was computed assuming a mean bulk density of  $3,000 \text{ kg m}^{-3}$  and average impact velocity of  $20.3 \text{ km s}^{-1}$ . The intrinsic impact frequency for these telescopic data was found using an average probability of impact for near-Earth asteroids of  $2 \times 10^{-9}$  per year for the entire population of asteroids. Lunar crater counts converted to equivalent impactor flux and assuming a geometric albedo of 0.25 (grey solid line) are shown for comparison<sup>9</sup>, although we note that contamination by secondary craters and modern estimates of the near-Earth-asteroid population that suggest lower albedos will tend to shift this curve to the right and downwards. Finally, we show the estimated influx from global airwave measurements conducted from 1960 to 1974, which detected larger (5–20 m) bolide impactors (red triangles)<sup>18</sup> using an improved method for energy estimation compared to earlier interpretations of the same data.

small number statistics. This limits our ability to validate the simulated CTH overpressure spatially.

Using our best estimate for the Chelyabinsk airburst energy, of about 500 kt, we have estimated the bolide flux at the Earth over the period from 1994 to mid-2013. This estimate is based on 20 years of total global coverage by the US government or infrasound sensors, more than doubling the earlier time coverage<sup>16,17</sup>. All events with estimated yields in excess of 1 kt are included. Figure 3 shows that this bolide flux at small sizes (less than 5 m in diameter) is in agreement within uncertainties with telescopic<sup>8</sup> data and earlier infrasonic<sup>18</sup> influx estimates. However, at larger diameters (15–30 m), both the bolide and infrasound<sup>18</sup> flux curves show an apparent impact rate at the Earth an order of magnitude larger than either that estimated by

telescopic surveys or the longer-term average impact rate provided by lunar cratering. In both cases these deviations well above the constant power-law slope of ref. 16 are due to single large events, so caution must be exercised owing to the small number statistics. A best-fit regression line to the bolide flux is given by  $N = aE^{-b}$ , where  $N$  is the cumulative number of objects with energy  $E$  (in kilotons) or more that impact the Earth per year,  $a = 3.31 \pm 0.11$  and  $b = -0.68 \pm 0.06$ . We note that excluding the rightmost two points in Fig. 3 (representing Chelyabinsk and two other events larger than 30 kt) produces a nearly identical slope.

Using the telescopic impact frequency<sup>8</sup> (green squares in Fig. 3) as a baseline for the 20-year period of the bolide survey, there is only a 13% chance that any random 20-year period would have an airburst as large or larger than Chelyabinsk. The independent 14-year survey by infrasound<sup>18</sup> (1960–74) detected a probable ~1.5-Mt airburst on 3 August 1963. Such a large event would be expected at the ~3% level during such a survey period. Although these deviations may be attributable to small number statistics, we note that Tunguska, with a source energy (energy released at the location of the explosion) of the order of 3–15 Mt (refs 15,19; shown as a horizontal line in Fig. 3) is also an extreme outlier (expected at the 2–10% level to have occurred during the past century). These events, taken together with Chelyabinsk, are increasingly suggestive of non-equilibrium in the impactor flux for near-Earth asteroids that are 10–50 m in diameter. This is manifested as a change in the power-law energy–frequency distribution at these sizes, similar to changes in the power law at other sizes<sup>20</sup>. This is also consistent with the recent origin of Chelyabinsk as a single near-Earth asteroid and a possible link to asteroid 86039 (ref. 10). Our findings support earlier interpretations of an influx maximum at this size range<sup>21,22</sup>. We note that telescopic surveys have only discovered about 500 near-Earth asteroids that are 10–20 m in diameter<sup>8</sup> (comparable to the Chelyabinsk asteroid) of an estimated near-Earth asteroid population (<http://ssd.jpl.nasa.gov>) of around  $2 \times 10^7$ , implying that a non-equilibrium impactor population at these sizes could be present but not yet apparent in the discovered near-Earth asteroid population.

Received 27 June; accepted 3 October 2013.

Published online 6 November 2013.

- Boslough, M. B. & Crawford, D. A. Low-altitude airbursts and the impact threat. *Int. J. Impact Eng.* **35**, 1441–1448 (2008).
- Harris, A. W. What Spaceguard did. *Nature* **453**, 1178–1179 (2008).
- Chapman, C. R. & Morrison, D. Impacts on the Earth by asteroids and comets: assessing the hazard. *Nature* **367**, 33–40 (1994).
- Collins, G. S., Melosh, H. J. & Marcus, R. A. Earth Impact Effects Program: a web-based computer program for calculating the regional environmental consequences of a meteoroid impact on Earth. *Meteorit. Planet. Sci.* **40**, 817–840 (2005).
- Chyba, C. F., Thomas, P. J. & Zahnle, K. J. The 1908 Tunguska explosion: atmospheric disruption of a stony asteroid. *Nature* **361**, 40–44 (1993).
- Hills, J. G. & Goda, M. P. The fragmentation of small asteroids in the atmosphere. *Astron. J.* **105**, 1114–1144 (1993).
- Glasstone, S. & Dolan, P. J. *The Effects of Nuclear Weapons* 3rd edn, 100–105 (US Gov. Printing Office, 1977).
- Harris, A. The value of enhanced NEO surveys. IAA-PDC13-05-09 (Planetary Defence Conference, IAA, 2013).
- Werner, S. C., Harris, A. W., Neukum, G. & Ivanov, B. A. The near-Earth asteroid size-frequency distribution: a snapshot of the lunar impactor size-frequency distribution. *Icarus* **156**, 287–290 (2002).
- Borovička, J. *et al.* The trajectory, structure and origin of the Chelyabinsk asteroidal impactor. *Nature* <http://dx.doi.org/10.1038/nature12671> (this issue).
- Brown, P. G. *et al.* Analysis of a crater-forming meteorite impact in Peru. *J. Geophys. Res.* **113**, 1–13 (2008).
- Toon, O. B., Zahnle, K. J., Morrison, D., Turco, R. P. & Covey, C. Environmental perturbations caused by the impacts of asteroids and comets. *Rev. Geophys.* **35**, 41–78 (1997).
- Revelle, D. O. On meteor-generated infrasound. *J. Geophys. Res.* **81**, 1217–1230 (1976).
- ReVelle, D. O. in *Proc. Int. Astron. Union Symp. No. 236 (NEOs, our Celestial Neighbors: Opportunity and Risk)* (eds Milani, A., Valsecchi, G. B. & Vokrouhlický, D.) 95–106 (Cambridge Univ. Press, 2007).
- Boslough, M. B. & Crawford, D. Shoemaker-Levy 9 and plume-forming collisions on Earth. *Ann. NY Acad. Sci.* **822**, 236–282 (1997).
- Brown, P. G., Spalding, R. E. & ReVelle, D. O. Tagliaferri, E. & Worden, S. P. The flux of small near-Earth objects colliding with the Earth. *Nature* **420**, 294–296 (2002).
- Ens, T. A., Brown, P. G., Edwards, W. N. & Silber, E. Infrasound production by bolides: a global statistical study. *J. Atmos. Sol. Terr. Phys.* **80**, 208–229 (2012).
- Silber, E. & ReVelle, D. O. Brown, P. G. & Edwards, W. N. An estimate of the terrestrial influx of large meteoroids from infrasonic measurements. *J. Geophys. Res.* **114**, E08006 (2009).
- Ben-Menahem, A. Source parameters of the Siberian explosion of June 30, 1908, from analysis and synthesis of seismic signals at four stations. *Phys. Earth Planet. Inter.* **11**, 1–35 (1975).
- Bland, P. A. & Artemieva, N. The rate of small impacts on Earth. *Meteorit. Planet. Sci.* **41**, 607–631 (2006).
- Cepelcha, Z. Influx of interplanetary bodies onto Earth. *Astron. Astrophys.* **263**, 361–366 (1992).
- Ward, S. & Asphaug, E. Asteroid impact tsunami: a probabilistic hazard assessment. *Icarus* **145**, 64–78 (2000).
- Brachet, N. *et al.* in *Infrasound Monitoring for Atmospheric Studies* (eds Le Pichon, A., Blanc, E. & Hauchecorne, A.) 77–118 (Springer, 2010).
- Brown, P. G. *et al.* The fall of the Grimsby meteorite—I: Fireball dynamics and orbit from radar, video, and infrasound records. *Meteorit. Planet. Sci.* **46**, 339–363 (2011).
- Borovička, J. *et al.* The Košice meteorite fall: atmospheric trajectory, fragmentation, and orbit. *Meteorit. Planet. Sci.* <http://dx.doi.org/10.1111/maps.12078> (2013).
- Tagliaferri, E., Spalding, R., Jacobs, C., Worden, S. P. & Erlich, A. in *Hazards Due to Comets and Asteroids* (ed. Gehrels, T.) 199–221 (Univ. Arizona Press, 1994).
- Rabinowitz, D., Helin, E., Lawrence, K. & Pravdo, S. A reduced estimate of the number of kilometer-sized near-Earth asteroids. *Nature* **403**, 165–166 (2000).

Supplementary Information is available in the online version of the paper.

**Acknowledgements** Funding was provided by the NASA co-operative agreement NNX11AB76A and the Czech institutional project RVO:67985815. D.P.D. acknowledges support from the Office of Naval Research. We appreciate discussions with F. Gilbert (of UCSD), J. Stevens (of SAIC), P. Earle and J. Bellini (of USGS). D. Dearborn provided assistance with video reductions.

**Author Contributions** P.G.B., N.B., D.B., L.C., W.E., L.G.E., M.G., A.L.P., J.D.A., P.M., P. Smets and R.W. performed various aspects of the identification, measurement and interpretation of infrasound records. L.A., C.d.G.-H., M.H. and G.L. collected and identified the airburst signals in seismic recordings as well as analysing and interpreting the seismic data. P.G.B., R.B., J.B., W.C., J.G., A.K., D.E.M., R.W., A.S. and P. Spurny helped in identifying important videos and their geolocation and various aspects of their measurements. M.B.B. and M.C.-B. performed bolide entry modelling. D.P.D. provided atmospheric model data and interpretation. Z.K., J.G. and R.J.W. performed video lightcurve analysis and calibrations and helped with their interpretation as well as performing measurements of video dust cloud features. R.E.S. and E.T. facilitated and interpreted US Government Sensor data. D.U. performed window breakage analysis. P.G.B. and E.S. performed analysis of acoustic propagation and associated computer code development. P.G.B. wrote the manuscript. All authors commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.G.B. (pbrown@uwo.ca).

# Stabilizing the magnetic moment of single holmium atoms by symmetry

Toshio Miyamachi<sup>1,2</sup>, Tobias Schuh<sup>1</sup>, Tobias Märkl<sup>1</sup>, Christopher Bresch<sup>1</sup>, Timofey Balashov<sup>1</sup>, Alexander Stöhr<sup>1,3</sup>, Christian Karlewski<sup>4</sup>, Stephan André<sup>4</sup>, Michael Marthaler<sup>4</sup>, Martin Hoffmann<sup>5,6</sup>, Matthias Geilhufe<sup>5</sup>, Sergey Ostanin<sup>5</sup>, Wolfram Hergert<sup>6</sup>, Ingrid Mertig<sup>5,6</sup>, Gerd Schön<sup>4</sup>, Arthur Ernst<sup>5,7</sup> & Wulf Wulfhekel<sup>1</sup>

Single magnetic atoms, and assemblies of such atoms, on non-magnetic surfaces have recently attracted attention owing to their potential use in high-density magnetic data storage and as a platform for quantum computing<sup>1–8</sup>. A fundamental problem resulting from their quantum mechanical nature is that the localized magnetic moments of these atoms are easily destabilized by interactions with electrons, nuclear spins and lattice vibrations of the substrate<sup>3–5</sup>. Even when large magnetic fields are applied to stabilize the magnetic moment, the observed lifetimes remain rather short<sup>5,6</sup> (less than a microsecond). Several routes for stabilizing the magnetic moment against fluctuations have been suggested, such as using thin insulating layers between the magnetic atom and the substrate to suppress the interactions with the substrate's conduction electrons<sup>2,3,5</sup>, or coupling several magnetic moments together to reduce their quantum mechanical fluctuations<sup>7,8</sup>. Here we show that the magnetic moments of single holmium atoms on a highly conductive metallic substrate can reach lifetimes of the order of minutes. The necessary decoupling from the thermal bath of electrons, nuclear spins and lattice vibrations is achieved by a remarkable combination of several symmetries intrinsic to the system: time reversal symmetry, the internal symmetries of the total angular momentum and the point symmetry of the local environment of the magnetic atom.

In a free atom the net spin and orbital angular momenta couple to the total angular momentum  $J$ , resulting in an atomic magnetic moment. The continuous rotation symmetry means that the  $2J+1$  magnetic sub-levels  $J_z = -J, -J+1, \dots, +J$  are degenerate in the absence of a magnetic field  $B$ , as illustrated in Fig. 1a. If the atom is

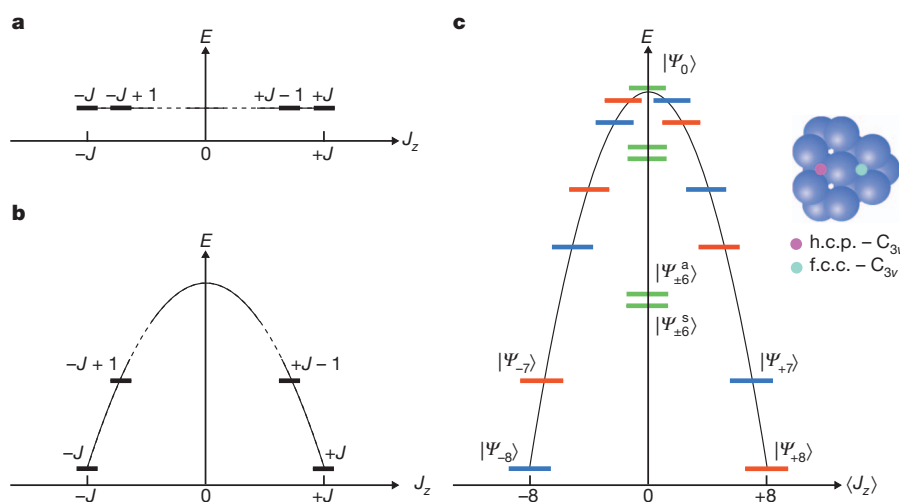
prepared in a specific magnetic state, this state is not stable against the slightest interaction of  $J$  with the environment, for example, with a thermal bath of electrons or nuclear spins, leading—without an energy cost—to transitions to other magnetic states and, hence, to a loss of orientation of the magnetic moment. Even the measurement of the magnetic state represents such an interaction.

The magnetic moment of an atom in a memory cell or quantum device can be stabilized by placing the atom on a substrate where the electrostatic interaction with the neighbouring atoms gives rise to magnetic anisotropy. This crystal field splits the formerly degenerate magnetic states. A decade ago, a giant magnetic anisotropy was discovered for single Co atoms on the Pt(111) surface<sup>1</sup>. Because of the broken inversion symmetry, a magnetic moment oriented normal to the surface plane is preferred over an in-plane orientation (the normal orientation is 9.3 meV lower in energy). To lowest order, this uniaxial anisotropy can be described by a simple crystal-field Hamiltonian  $\mathcal{H} = DJ_z^2$ , where  $D$  is the uniaxial anisotropy constant. Figure 1b depicts the resulting states in a shape of a down-turned parabola ( $D < 0$ ) with the two degenerate ground states  $|+J\rangle$  and  $|-J\rangle$ .

The main mechanism for magnetization reversal are spin-flip events mediated by the exchange interaction with substrate electrons. The interaction operator can be written as:

$$\mathcal{V} = \mathbf{J} \cdot \boldsymbol{\sigma} = J_z \sigma_z + \frac{1}{2} (J_+ \sigma_- + J_- \sigma_+), \quad (1)$$

where  $\boldsymbol{\sigma}$  is the spin of the scattering electron and  $J_+$  and  $J_-$  are the ladder operators within the  $J$ -multiplet ( $\sigma_+$  and  $\sigma_-$  are the analogous ladder operators for the spin)<sup>3,9,10</sup>. A spin-flip scattering event changes



**Figure 1 | Magnetic sublevels of atoms with a total angular momentum  $J$ .** a, Degenerate sublevels of a free atom. b, Sublevels split by uniaxial anisotropy. c, The states mix in the presence of a crystal-field Hamiltonian of  $C_{3v}$  symmetry. The degenerate green states split into symmetric (s) and antisymmetric (a) states. The inset shows the f.c.c. and h.c.p. adsorption sites.

<sup>1</sup>Physikalisches Institut, Karlsruhe Institute of Technology (KIT), Wolfgang-Gaede-Straße 1, 76131 Karlsruhe, Germany. <sup>2</sup>Institute of Solid State Physics, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwashi, Chiba 277-8581, Japan. <sup>3</sup>Max-Planck-Institut für Festkörperforschung, Heisenbergstraße 1, 70569 Stuttgart, Germany. <sup>4</sup>Institut für Theoretische Festkörperforschung, Karlsruhe Institute of Technology (KIT), Wolfgang-Gaede-Straße 1, 76131 Karlsruhe, Germany. <sup>5</sup>Max-Planck-Institut für Mikrostrukturphysik, Weinberg 2, 06120 Halle, Germany. <sup>6</sup>Institut für Physik, Martin-Luther-Universität Halle-Wittenberg, 06099 Halle, Germany. <sup>7</sup>Wilhelm-Ostwald-Institut für Physikalische und Theoretische Chemie, Universität Leipzig, Linnéstraße 2, 04103 Leipzig, Germany.



$J_z$  by  $\pm\hbar$ . From this, one could conjecture that a complete reversal over the full barrier is strongly suppressed because it requires at least  $2J$  scattering events and the energy barrier is large compared to thermal energies at cryogenic temperatures (for example, 0.34 meV at 4 K). This raised the hope of storing a magnetic bit in the two ground states.

However, in spite of the high magnetic anisotropy, the actual stability of the ground state is very limited owing to the strong hybridization of the  $3d$  states of the transition atom with the substrate electrons. This leads to lifetimes of the excited states of the order of  $\tau \approx 10$  fs (refs 4 and 6) and to an energy smearing of the excited states of the order of the barrier height due to the Heisenberg uncertainty principle. As a result, a reversal is possible without the need for thermal energy.

Longer lifetimes were achieved with atoms deposited on thin insulating films<sup>5</sup> where the interaction with the substrate electrons is weakened. In this case, a close inspection of the magnetic transitions induced by the tunnelling electrons of a scanning tunnelling microscope (STM) showed more transitions<sup>3</sup> than expected from the simple uniaxial Hamiltonian. In fact, the magnetic anisotropy of an adsorbed atom is hardly strictly uniaxial. Depending on the point symmetry of the adsorption site, additional terms enter the crystal-field Hamiltonian  $\mathcal{H}$ . In general, the crystal-field Hamiltonian can be written as:

$$\mathcal{H} = \sum_{n=0}^{\infty} \left( \sum_{m=0}^n B_n^m O_n^m + \sum_{m=1}^n \tilde{B}_n^m \tilde{O}_n^m \right), \quad (2)$$

where  $B_n^m$  and  $\tilde{B}_n^m$  are the anisotropy constants and  $O_n^m$  and  $\tilde{O}_n^m$  are the crystal-field or Stevens operators. The latter are polynomials of order  $n$  in the operators  $J_z$ ,  $J_+$  and  $J_-$  with order  $m$  in the ladder operators<sup>11</sup> and induce a zero-field splitting in the multiplet. Owing to time-reversal symmetry, states with opposite magnetic moment have the same energy for  $\mathbf{B} = 0$ . Thus, all terms with odd  $n$  are forbidden. All terms with  $m = 0$  only contain powers of  $J_z$  and thus do not mix the  $J_z$  eigenstates; for example,  $O_2^0$  is essentially the already-introduced operator of the uniaxial anisotropy. All terms with  $n > 2J$  vanish, because the operators act within the  $2J + 1$  states of the multiplet. Likewise, terms with  $n > 2l$  vanish, where  $l$  is the orbital angular momentum of the individual electrons in the incomplete shell of the atom. Finally, all operators vanish that are incompatible with the point symmetry of the adsorption site; for example, all  $\tilde{B}_n^m$  vanish for a mirror plane normal to the surface.

In the following, we concentrate on atoms adsorbed in three-fold hollow sites of the (111) surface of face-centred cubic (f.c.c.) crystals. Both the f.c.c. and the hexagonal close-packed (h.c.p.) adsorption sites have a three-fold symmetry with a vertical mirror plane (and its rotations) perpendicular to the surface (see inset of Fig. 1c), that is, a  $C_{3v}$  symmetry. Encouraged by a previous study<sup>12</sup>, we used rare earth atoms instead of transition metals, because the magnetic anisotropies of the former are typically larger than those of  $3d$  atoms, and their  $4f$  orbitals hybridize less with the substrate, resulting in potentially longer lifetimes. Thus, for rare earth atoms (with  $l = 3$ ) in zero magnetic field the Hamiltonian can be written as<sup>13</sup>:

$$\mathcal{H} = B_2^0 O_2^0 + B_4^0 O_4^0 + B_6^0 O_6^0 + B_4^3 O_4^3 + B_6^3 O_6^3 + B_6^6 O_6^6. \quad (3)$$

The operators  $O_n^m$  are given explicitly in the Supplementary Information. Although the first three terms do not mix the  $J_z$  eigenstates, the last three mix  $J_z$  states differing in  $J_z$  by 3 or 6.

The energy spectrum of the Hamiltonian (equation (3)) for a holmium (Ho) impurity ( $J = 8$ , with ten  $4f$  electrons<sup>14</sup>) and dominating, negative  $B_2^0$  is illustrated in Fig. 1c. Overall, the parabolic energy dependence is preserved when the states are plotted according to their expectation value of  $J_z$ . Every eigenstate is a mixture of several  $|J_z\rangle$  states differing by  $\Delta J_z = \pm 3$  and thus three classes of eigenstates form, marked in Fig. 1c by three different colours. Of the states marked in green, the  $|\pm 3\rangle$  and  $|\pm 6\rangle$  states are pairwise degenerate when considering only the first three operators in equation (3). They are, however, strongly mixed by the last three operators, resulting in new non-magnetic, energy-split

states of different symmetry ('s' for symmetric and 'a' for antisymmetric). The red and the blue states form degenerate doublets.

In general, depending on  $J$ , three cases can arise. First, the two ground states may be of the same colour and get strongly mixed and split. This scenario is often called the tunnelling case. The eigenstates are not pure in  $J_z$  so the preparation of a pure state leads to tunnelling to a state of reversed magnetization by the Stevens operators alone, and consequently to short relaxation times<sup>5</sup>. Second, the two ground states  $\Psi_{-J}$  and  $\Psi_{+J}$  do not mix, but are linked by spin-flip transitions due to  $\mathcal{V}$  (see equation (1)). This is the so-called Kondo case where scattering with a single electron can lead to transitions between the two ground states with zero energy cost. Again, lifetimes are short. At low temperatures the conduction electrons can even form a Kondo singlet with the atom. Third, the transition between the two ground states by scattering with a single electron is forbidden (that is, the matrix element of  $\mathcal{V}$  vanishes). Thus, the magnetic moment of the atom is decoupled from the electrons of the substrate to first order, and reversal is possible only by sequential scattering via intermediate states of higher energy. Depending on the excitation energy, these processes can be suppressed at low temperatures, effectively restoring the barrier and leading to long lifetimes.

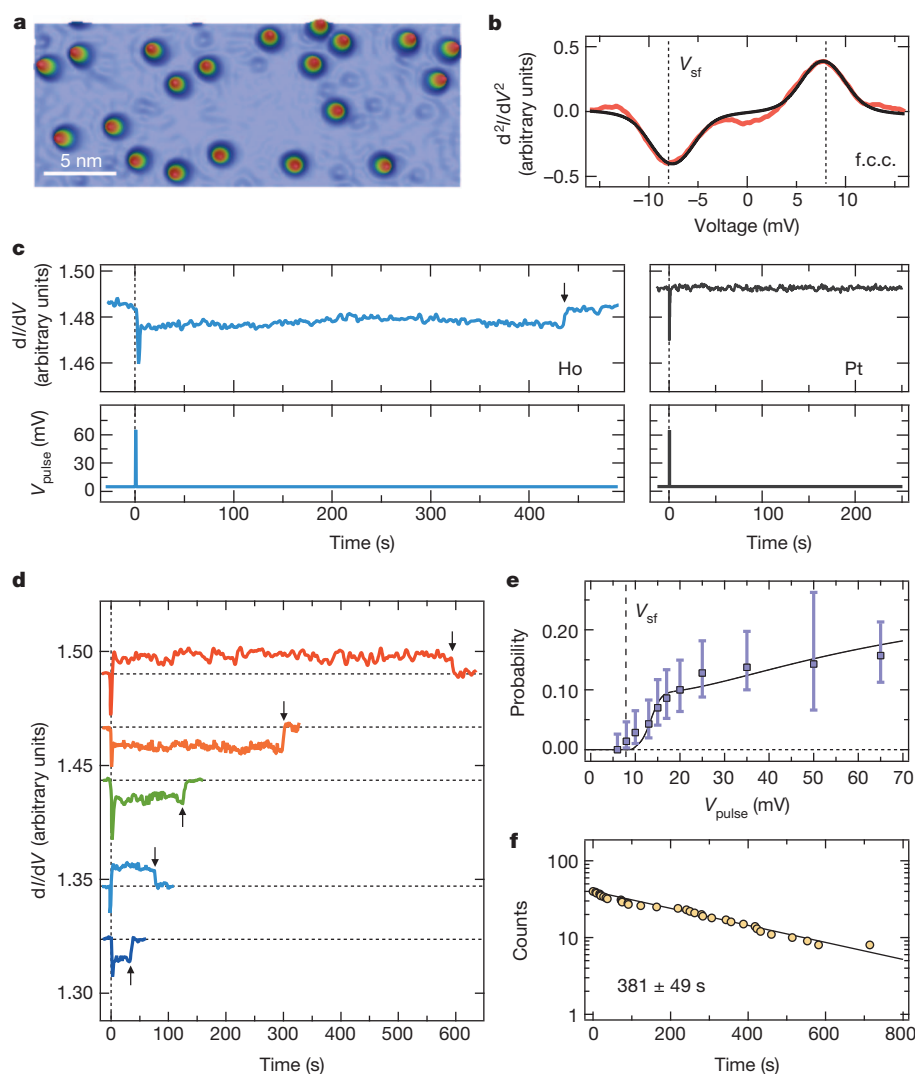
The last scenario is realized for Ho on Pt(111), that is,  $\langle \Psi_{-8} | \mathcal{V} | \Psi_{+8} \rangle = 0$ . This is a consequence of the combination of the symmetries and the value of  $J$  and follows from group-theoretical arguments. To show this, we investigate how the matrix element  $\langle \Psi_{-8} | J_i | \Psi_{+8} \rangle$ , where  $i \in \{x, y, z\}$ , behaves under time reversal. For  $\mathbf{B} = 0$ ,  $\mathcal{H}$  commutes with the time-reversal operator  $\mathcal{T}$ , and  $\mathcal{T} J_i = -J_i \mathcal{T}$ . Further,  $\mathcal{T}^2 = 1$  because the number of electrons in Ho is even,  $\mathcal{T} | \Psi_{\pm 8} \rangle = | \Psi_{\mp 8} \rangle$  (see Supplementary Information) and  $\mathcal{T}$  is anti-unitary, that is,  $\langle \chi | \phi \rangle = \langle \mathcal{T} \phi | \mathcal{T} \chi \rangle = \langle \mathcal{T} \chi | \mathcal{T} \phi \rangle^*$ , where the asterisk indicates the complex conjugate. Thus we can write:

$$\begin{aligned} \langle \Psi_{-8} | \overbrace{J_i}^{|\phi\rangle} | \Psi_{+8} \rangle &= \langle \mathcal{T} \Psi_{-8} | \mathcal{T} J_i | \Psi_{+8} \rangle^* \\ &= -\langle \Psi_{+8} | J_i | \mathcal{T} \Psi_{+8} \rangle^* = -\langle \Psi_{+8} | J_i | \Psi_{-8} \rangle^* \\ &= -\langle \Psi_{-8} | J_i | \Psi_{+8} \rangle, \quad i \in \{x, y, z\} \end{aligned} \quad (4)$$

Thus, the matrix elements of  $J$  vanish as a consequence of symmetries and direct transitions between the ground states cannot be induced by scattering electrons. We expect long lifetimes at low temperatures.

In the experiments, Ho atoms were deposited on a clean Pt(111) substrate at 4.4 K, leading to individual Ho atoms on the surface (adatoms) as seen by low-temperature STM (compare Fig. 2a). First, a non-magnetic tungsten tip was used to image the surface and investigate excitations of the atom. The second derivative of the tunnelling current  $I$  was measured as function of the bias voltage  $V$ , while positioning the tip over the Ho atoms, as shown in Fig. 2b. The  $d^2I/dV^2$  spectra taken at 4.4 K reveal the typical dip–peak structure at a bias voltage  $V_{sf}$  as seen in many other single-atom systems<sup>4,6,12</sup>. Peak and dip are well separated, speaking for an inelastic excitation as opposed to the Kondo effect. As shown later by spin-polarized STM measurements, this excitation is magnetic and corresponds to a spin-flip transition from the ground state to the first excited state. The analysis of the spectra of many atoms reveals that the excitation energy of atoms in f.c.c. sites (about 8 meV) is higher than in h.c.p. sites (about 6 meV) and that the adsorption energy of f.c.c. sites is higher (see Supplementary Information).

The anisotropy constants can also be obtained from first principles, by applying a first-order perturbation theory<sup>15</sup> (see the Supplementary Information for more details). These calculations confirm that Ho on Pt(111) is indeed in a  $J = 8$  state and shows an out-of plane easy axis. For f.c.c. Ho atoms, the first excited magnetic state is 7.7 meV above the ground state, in good agreement with the experimental observation.



**Figure 2 | Magnetic behaviour of Ho atoms ( $J = 8$ ) adsorbed on Pt(111).** **a**, Three-dimensional topographic STM image of single Ho atoms adsorbed on Pt(111) at 4.4 K. **b**, Inelastic tunnelling spectrum ( $d^2I/dV^2$ ) recorded on top of a single Ho atom in the f.c.c. position (red), showing an inelastic excitation at  $V_{sf} \approx 8$  meV, as determined by a fit to the data (black). **c**, Time traces of the  $dI/dV$  signal recorded with a spin-polarized tip on a Ho atom in the f.c.c. position and on bare Pt ( $V = 5$  mV,  $I = 1$  nA). At  $t = 0$  a voltage pulse  $V_{pulse}$  above the inelastic excitation threshold was applied. Although on bare Pt(111) no changes of the  $dI/dV$  signal were observed, the pulse induced a magnetic transition between the two ground states of Ho, as deduced from the observed change of the spin-polarized  $dI/dV$  signal. After some delay, the atoms spontaneously switch back to the initial state, as seen by a second change of the signal (indicated by arrow). **d**, A series of switching measurements of Ho atoms representing a wide range of the decay times indicated by arrows. **e**, Magnetic switching probabilities and their standard deviation as function of the pulse height as well as a fit (black line) to the data obtained from a master equation for the induced transitions. **f**, Decay of 40 states prepared by pulses (dots) and an exponential fit (black line).

In the following we address atoms on the high-anisotropy f.c.c. site only. The large magnetic excitation energy would lead to a stable magnetic moment at low temperatures, because we expect from theory that scattering with the electron bath cannot directly switch the magnetization of the atom between the two ground states. Only if a scattering electron (from the tunnelling current, for example) has sufficient energy can it bring the atom from the ground state to the first excited state, which in turn can relax to either of the two ground states via a second inelastic scattering event with substrate electrons.

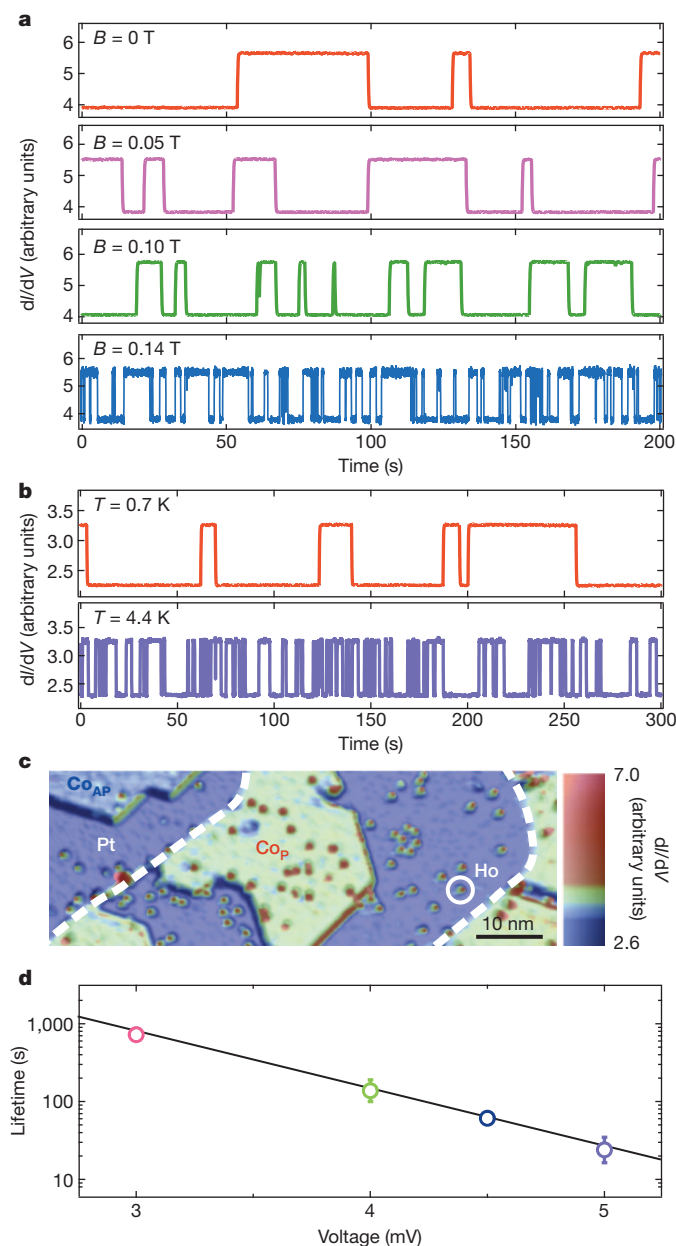
To investigate this scenario, we carried out spin-polarized STM measurements with anti-ferromagnetic tips that have a spin polarization perpendicular to the surface<sup>16</sup>. We stabilized the tip over a Ho atom at  $I = 1$  nA and  $V = 5$  mV, which is below the excitation threshold. Thus, the tunnelling electrons cannot switch the atom but probe the spin polarization on top of the atom. The  $dI/dV$  signal depends on the relative orientation of the tip magnetic moment and the Ho moment. Then we applied 0.2-ms voltage pulses, while the feedback was turned off. If the atom switches between the two ground states during the pulse, the  $dI/dV$  signal is different before and after the pulse. We varied the height of the pulse up to 65 mV.

Figure 2c shows examples of recorded time traces of the  $dI/dV$  signal at 1.1 K in cases where we observed a switching and, for comparison, on bare Pt(111). The spin signal recorded above the atom has two distinct values representing the two magnetic ground states with the magnetization perpendicular to the surface, whereas the signal on Pt(111)

is not influenced by the pulse, thus excluding tip changes. After switching the magnetic moment of the Ho atom, it stays in the reversed state for times in the range from seconds to minutes until it spontaneously switches back, as indicated by the arrows in Fig. 2d. This illustrates long lifetimes, as will be discussed below.

By applying many pulses to an individual atom, we determined the probability that a voltage pulse leads to a reversal of the magnetic state (see Fig. 2e). In agreement with the inelastic spectrum, we observe a threshold for reversal at about 8 mV. Above this, the switching probability per pulse increases continuously owing to the increase of the tunnelling current with pulse voltage, because the feedback loop is opened during the pulse. The switching probability depends on the branching ratio of the relaxation from the excited states to either of the ground states, so the switching probability contains information about the anisotropy constants. The solid line in Fig. 2e was calculated using a master equation of the dynamics of the states' population with anisotropy constants determined from first principles and fitting the coupling strength of the Ho spin to the substrate and tunnelling electrons (for details see the Supplementary Information). The switching probabilities for higher-voltage pulses depend strongly on the tip spin polarization, which can be determined from the fit to be about 20%. The curve fits the experimental data excellently, confirming the quantum mechanical nature of the Ho spin.

To investigate the stability of the ground states, we repeated the switching experiment 40 times and plotted the observed population



**Figure 3 | Lifetimes of adsorbed Ho atoms as function of external parameters.** **a**, Spontaneous transitions between the two ground states for different magnetic fields along the surface normal observed with spin-polarized  $dI/dV$  signal recorded at  $0.7$  K,  $5$  mV and  $50$  nA. **b**, Spontaneous transitions between the two ground states for two different temperatures without applied magnetic field recorded at  $5$  mV and  $50$  nA. **c**, Spin-polarized  $dI/dV$  map of Ho atoms and Co islands on Pt(111) ( $V = 300$  mV,  $I = 1.5$  nA). Two distinct signals are observed on Co islands magnetized parallel (P) and antiparallel (AP) to the tip. Positions of Pt atomic steps are indicated by dashed lines. **d**, Lifetimes extracted from spontaneous transitions at  $0.7$  K without applied magnetic field and  $1$  nA as function of sample bias on the marked Ho atom in **c**. Error bars represent the standard deviation of a fit to an exponential decay (28, 23, 102 and 94 transitions for  $3.0$ -mV,  $4.0$ -mV,  $4.5$ -mV and  $5.0$ -mV sample biases, respectively).

decay (see Fig. 2f). The data follow a simple exponential decay with a lifetime  $\tau$  of  $381 \pm 49$  s, indicating an efficient decoupling of the Ho spin from the conduction electrons.

To understand the value of  $\tau$ , we turn back to the original explanation: direct transitions between the ground states by tunnelling and single spin-flip events are forbidden due to the time-reversal symmetry. Thus, transitions between the ground states can only be achieved

by excitation to the first (or higher) excited state by substrate or tip electrons in the high energy tail of the thermal distribution or by breaking time-reversal symmetry using a magnetic field. To verify this, we investigated the dependence of  $\tau$  on bias voltage, temperature and magnetic field.

Figure 3a illustrates the effect of a magnetic field applied perpendicularly to the surface at a temperature of  $0.7$  K. Similarly to the previous experiments, the spin-polarized  $dI/dV$  signal on top of a Ho atom was recorded as a function of time. To increase the signal-to-noise ratio, a larger tunnelling current of  $50$  nA was used while keeping  $V = 5$  mV. In this experiment, we did not apply pulses but observed spontaneous transitions between the two ground states. An increase of the switching frequency can be clearly observed with increasing magnetic field because spin-flip scattering between the two ground states is no longer forbidden (for details see the Supplementary Information). Similarly,  $\tau$  decreases with increasing temperature, as illustrated in Fig. 3b ( $V = 5$  mV,  $I = 50$  nA).

To study the bias dependence, a different sample was used. Figure 3c shows a spin-polarized image ( $dI/dV$  map) of adsorbed Ho atoms. To verify the spin-polarization of the tip, Co islands were deposited that decorate the step edges of the Pt(111) substrate. The Co islands are magnetized out-of-plane<sup>17</sup> and the contrast between the two islands indicates that the tip is sensitive to the out-of-plane magnetization. Figure 3d shows  $\tau$  as a function of  $V$  ( $I = 1$  nA,  $T = 0.7$  K), obtained by analysis of spontaneous transitions. Even though the bias voltages are below the first excitation energy,  $\tau$  drops exponentially with rising voltage. This is due to the exponentially increasing number of thermal tunnelling electrons at higher bias voltages that have enough energy to excite the atom to the first excited state. Owing to the close proximity to the Co island, the Ho atom experiences a non-vanishing magnetic stray field of the island. This leads to a shorter  $\tau$  at  $5$  mV than in the first measurement in spite of the lower temperature. However, at a  $3$  mV bias voltage, the lifetime is much longer,  $\tau = 729 \pm 12$  s.

An interesting side effect of the forbidden spin-flip scattering by conduction electrons is that magnetic interactions between Ho atoms mediated by the Ruderman–Kittel–Kasuya–Yosida (RKKY) interaction<sup>18</sup> do not cause the exchange of angular momentum to lowest order, such that the observed  $\tau$  is nearly independent of the local arrangement of the Ho atoms, as discussed in the Supplementary Information.

In conclusion, general symmetry requirements open up a new pathway to stabilize the magnetic moment of adsorbed atoms and decouple it from conduction electrons and nuclear spins. This method could potentially be used in data storage and quantum computing. By applying a magnetic field, the interactions could be switched back on, thus allowing controlled quantum manipulation of the spins.

## METHODS SUMMARY

The Pt(111) sample was cleaned by argon-ion sputtering followed by annealing to  $850$  K. The sample surfaces were checked for impurities by STM at  $4.4$  K before deposition of Ho. Ho atoms were deposited onto the clean surfaces at  $4.4$  K directly in the STM. Unpolarized STM tips were prepared from tungsten wires and were cleaned *in situ* by flashing to above  $2,500$  K. Spin-polarized tips were prepared by depositing about 20 monolayers of Mn or Cr *in situ* onto the tip apex followed by gentle annealing. Inelastic  $d^2I/dV^2$  spectra were recorded using a modulation frequency of about  $16$  kHz at a root-mean-square voltage of  $2.4$  mV and detecting the second harmonics with a lock-in technique. Spin-polarized STM measurements were performed recording the  $dI/dV$  signal with coated tips using a modulation of  $720$  Hz and  $1$  mV root-mean-square voltage. Voltage pulses were given with the feedback loop open to avoid destabilizing the tip position. Tip changes due to voltage pulses were excluded by checking between every measurement on Ho atoms also the time trace of pulses given on bare Pt(111). For tests of the out-of-plane sensitivity of the spin-polarized tips, one monolayer high Co islands were grown on Pt(111) at  $360$  K before inserting the sample into the STM. All experiments were performed using a home-built cryogenic Joule–Thomson cooled STM<sup>19</sup>.

First-principles calculations were carried out within the density functional theory. The crystalline structure was determined with the Vienna Ab-initio Simulation Package (VASP) method, well known for providing precise total energy and



forces<sup>20,21</sup>. The obtained structural information was further used for calculations of the electronic and magnetic properties of single Ho atoms on the Pt(111) surface using a self-consistent relativistic full-charge Green's function method, which is specially designed for semi-infinite systems and embedded real-space clusters<sup>22,23</sup>. A self-interaction correction method and a local density approximation (LDA) approximation including Hubbard  $U$  corrections were applied to provide an adequate description of strongly localized Ho  $f$  electrons<sup>24,25</sup>.

Received 15 August; accepted 9 October 2013.

- Gambardella, P. *et al.* Giant magnetic anisotropy of single cobalt atoms and nanoparticles. *Science* **300**, 1130–1133 (2003).
- Heinrich, A. J., Gupta, J. A., Lutz, C. P. & Eigler, D. M. Single-atom spin-flip spectroscopy. *Science* **306**, 466–469 (2004).
- Hirjibehedin, C. F. *et al.* Large magnetic anisotropy of a single atomic spin embedded in a surface molecular network. *Science* **317**, 1199–1203 (2007).
- Balashov, T. *et al.* Magnetic anisotropy and magnetization dynamics of individual atoms and clusters of Fe and Co on Pt(111). *Phys. Rev. Lett.* **102**, 257203 (2009).
- Loth, S., Etzkorn, M., Lutz, C. P., Eigler, D. M. & Heinrich, A. J. Measurement of fast electron spin relaxation times with atomic resolution. *Science* **329**, 1628–1630 (2010).
- Khajetoorians, A. A. *et al.* Itinerant nature of atom-magnetization excitation by tunneling electrons. *Phys. Rev. Lett.* **106**, 037205 (2011).
- Loth, S., Baumann, S., Lutz, C. P., Eigler, D. M. & Heinrich, A. J. Bistability in atomic-scale antiferromagnets. *Science* **335**, 196–199 (2012).
- Khajetoorians, A. A. *et al.* Current-driven spin dynamics of artificially constructed quantum magnets. *Science* **339**, 55–59 (2013).
- Fransson, J. Spin inelastic electron tunneling spectroscopy on local spin adsorbed on surface. *Nano Lett.* **9**, 2414–2417 (2009).
- Schuh, T. *et al.* Magnetic anisotropy and magnetic excitations in supported atoms. *Phys. Rev. B* **84**, 104401 (2011).
- Bleaney, B. & Stevens, K. W. H. Paramagnetic resonance. *Rep. Prog. Phys.* **16**, 108–159 (1953).
- Schuh, T. *et al.* Magnetic excitations of rare earth atoms and clusters on metallic surfaces. *Nano Lett.* **12**, 4805–4809 (2012).
- Wybourne, B. G. *Spectroscopic Properties of Rare Earths* Ch. 4.4, 166 (Wiley, 1965).
- Coe, J. M. D. *Magnetism and Magnetic Materials* 114 (Cambridge Univ. Press, 2009).
- Richter, M., Oppeneer, P., Eschrig, H. & Johansson, B. Calculated crystal-field parameters of SmCo<sub>5</sub>. *Phys. Rev. B* **46**, 13919–13927 (1992).
- Wiesendanger, R. Spin mapping at the nanoscale and atomic scale. *Rev. Mod. Phys.* **81**, 1495–1550 (2009).
- Rusponi, S. *et al.* The remarkable difference between surface and step atoms in the magnetic anisotropy of two-dimensional nanostructures. *Nature Mater.* **2**, 546–551 (2003).
- Zhou, L. *et al.* Strength and directionality of surface Ruderman-Kittel-Kasuya-Yosida interaction mapped on the atomic scale. *Nature Phys.* **6**, 187–191 (2010).
- Zhang, L., Miyamachi, T., Tomanić, T., Dehm, R. & Wulfhekel, W. A compact sub-Kelvin ultrahigh vacuum scanning tunneling microscope with high energy resolution and high stability. *Rev. Sci. Instrum.* **82**, 103702 (2011).
- Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
- Hafner, J. Ab-initio simulations of materials using VASP: density-functional theory and beyond. *J. Comput. Chem.* **29**, 2044–2078 (2008).
- Lüders, M., Ernst, A., Temmerman, W. M., Szotek, Z. & Durham, P. J. Ab initio angle-resolved photoemission in multiple-scattering formulation. *J. Phys. Condens. Matter* **13**, 8587–8606 (2001).
- Zeller, R. & Dederichs, P. H. Electronic structure of impurities in Cu, calculated self-consistently by Korringa-Kohn-Rostoker Green's-function method. *Phys. Rev. Lett.* **42**, 1713–1716 (1979).
- Perdew, J. P. & Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* **23**, 5048–5079 (1981).
- Anisimov, V. I., Zaanen, J. & Andersen, O. K. Band theory and Mott insulators: Hubbard  $U$  instead of Stoner  $I$ . *Phys. Rev. B* **44**, 943–954 (1991).

Supplementary Information is available in the online version of the paper.

**Acknowledgements** We acknowledge funding by the German Science Foundation (DFG) grant number Wu 349/4-2, the DFG priority programme SPP 1538 Spin Caloric Transport and the DFG Collaborative Research Centre SFB 762 Functionality of Oxide Interfaces. The calculations were performed at the Rechenzentrum Garching of the Max Planck Society.

**Author Contributions** W.W. conceived the experiments, and T. Miyamachi, T.S., T. Märkl, A.S. and C.B. carried them out. The data were analysed by T. Miyamachi, T.S., T. Märkl, C.B., T.B. and W.W. Group theory of the crystal field was performed by T.S., T.B., C.B. and W.W. Master equations were analysed by C.K., S.A., M.M. and G.S. Ab initio calculations were performed by M.H., M.G., S.O., W.H., I.M. and A.E. The manuscript was written by T.B. and W.W. Figures were prepared by T. Miyamachi. All authors discussed the results and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.W. ([wulf.wulfhekel@kit.edu](mailto:wulf.wulfhekel@kit.edu)).

# Guided hierarchical co-assembly of soft patchy nanoparticles

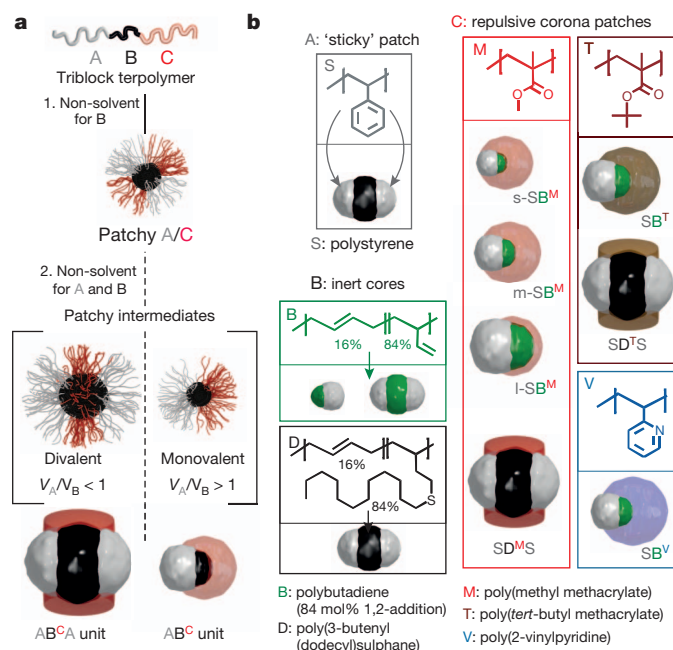
André H. Gröschel<sup>1†</sup>, Andreas Walther<sup>2</sup>, Tina I. Löbbling<sup>1</sup>, Felix H. Schacher<sup>3</sup>, Holger Schmalz<sup>1</sup> & Axel H. E. Müller<sup>1†</sup>

The concept of hierarchical bottom-up structuring commonly encountered in natural materials provides inspiration for the design of complex artificial materials with advanced functionalities<sup>1,2</sup>. Natural processes have achieved the orchestration of multicomponent systems across many length scales with very high precision<sup>3,4</sup>, but man-made self-assemblies still face obstacles in realizing well-defined hierarchical structures<sup>5–11</sup>. In particle-based self-assembly, the challenge is to program symmetries and periodicities of superstructures by providing monodisperse building blocks with suitable shape anisotropy or anisotropic interaction patterns ('patches'). Irregularities in particle architecture are intolerable because they generate defects that amplify throughout the hierarchical levels. For patchy microscopic hard colloids, this challenge has been approached by using top-down methods (such as metal shading or microcontact printing), enabling molecule-like directionality during aggregation<sup>12–16</sup>. However, both top-down procedures and particulate systems based on molecular assembly struggle to fabricate patchy particles controllably in the desired size regime (10–100 nm). Here we introduce the co-assembly of dynamic patchy nanoparticles—that is, soft patchy nanoparticles that are intrinsically self-assembled and monodisperse—as a modular approach for producing well-ordered binary and ternary supracolloidal hierarchical assemblies. We bridge up to three hierarchical levels by guiding triblock terpolymers (length scale ~10 nm) to form soft patchy nanoparticles (20–50 nm) of different symmetries that, in combination, co-assemble into substructured, compartmentalized materials (>10 µm) with predictable and tunable nanoscale periodicities. We establish how molecular control over polymer composition programs the building block symmetries and regulates particle positioning, offering a route to well-ordered mixed mesostructures of high complexity.

Until now, research on block copolymer self-assembly in solution focused mostly on multicompartment micelles or the crystallization-driven formation of compartmentalized structures bridging one hierarchical level<sup>17–21</sup>. Only a few studies have considered these nanoscale superstructures to be soft colloidal building blocks (CBBs) that can be self-assembled on higher levels<sup>22–26</sup>. The central challenge remains to devise patchy CBBs with sufficient precision qualifying as supraparticular tectons; that is, monodisperse in size and modified with defined repulsive and attractive surface patches providing directional interaction patterns<sup>27</sup>. Equally problematic while operating on the nanoscale are controlled particle positioning (interparticle forces may exceed particle size), the general lack of target-oriented and predictable self-assembly protocols, and convincing visualization of nano-separated multiphase organic materials. Responsive multiblock copolymers (1–10 nm) are potential candidates for the versatile bottom-up design (geometry, patchiness and dimension) of surface-compartmentalized nanoparticles (20–50 nm) with molecular precision (Fig. 1a). The particles themselves then transmit the information needed for defined higher-level co-assembly into mesoscale structures (0.1–10 µm) of controllable size and periodicity. In contrast with 'static', patchy microparticles (hard

spheres)<sup>14,27,28</sup>, these polymer-based nanoparticles are intrinsically self-assembled, soft and 'dynamic', offering the attractive feature of assembly or disassembly on demand—from molecules to CBBs and beyond. We recently approached the design of such a system by fabricating near-monodisperse, monovalent and divalent CBBs ('monomeric units') by the self-assembly of ABC triblock terpolymers in selective solvents<sup>23</sup>. Shaping building-block geometries and understanding both interparticle interactions and aggregation behaviour led us to the hypothesis that the concept of soft nanoparticle self-assembly could be extended to the rational design of supracolloidal co-assemblies by suitable combinations of building blocks. Mixed particle co-assemblies across multiple hierarchical levels open a new level of complexity and have yet to be addressed.

The general design criteria—developed in an experimental approach—for the guided co-assembly of multiple CBBs with distinct valences are few and simple, and should be widely applicable to polymer-decorated



**Figure 1 | Preparation and configuration of soft colloidal building blocks (CBBs).** **a**, Dispersion of ABC triblock terpolymers in a non-solvent for B yields B-core particles with A/C corona patches. During transfer into a non-solvent for A and B, these develop into monovalent  $AB^C$  and divalent  $AB^CA$  units with sticky A patches. **b**, CBBs feature associative ('sticky') polystyrene (S) patches reversibly made solvophobic on demand, chemically different (inert) polybutadiene (B) or poly(3-butenyl(dodecyl)sulfane) (D) core material physically holding the CBBs together, and poly(methyl methacrylate) (M), poly(*tert*-butyl methacrylate) (T) or poly(2-vinylpyridine) (V) as solubilizing/repulsive corona varying in polarity and functionality.

<sup>1</sup>Makromolekulare Chemie II, Universität Bayreuth, D-95440 Bayreuth, Germany. <sup>2</sup>DWI at RWTH Aachen University, Institute for Interactive Materials Research, 52056 Aachen, Germany. <sup>3</sup>Institut für Organische Chemie und Makromolekulare Chemie und Jena Center for Soft Matter, Friedrich Schiller Universität Jena, D-07743 Jena, Germany. <sup>†</sup>Present addresses: Department of Applied Physics, Aalto University, FI-02150 Espoo, Finland (A.H.G.); Institute of Organic Chemistry, Johannes Gutenberg-Universität, D-55099 Mainz, Germany (A.H.E.M.).

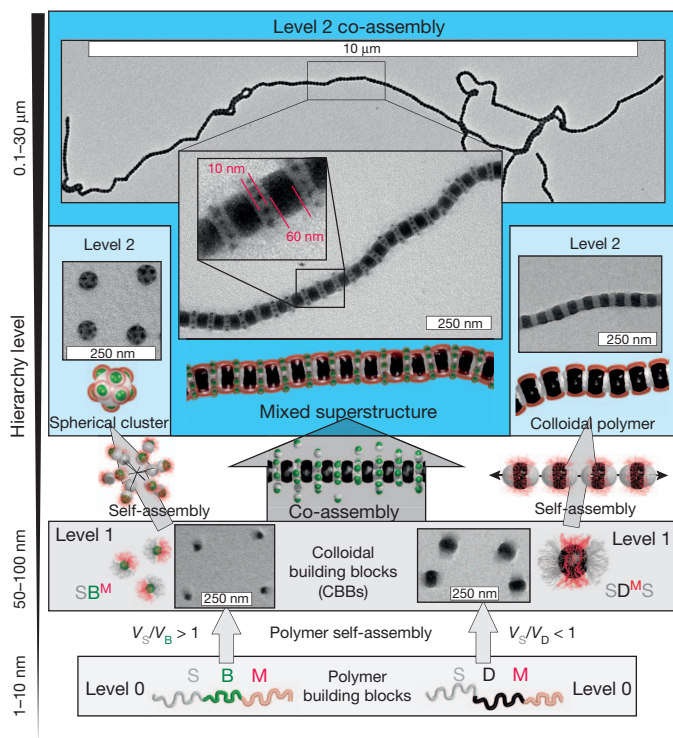
particles (Fig. 1a). One essential requirement is tunable and defined attractive interaction patterns responsive to solvent quality or other external stimuli to favour near-monodisperse structures on different length scales. Modern polymer synthesis provides us with well-defined block copolymers with a wide range of properties and responses<sup>29</sup>. We developed a set of ABC triblock terpolymers into which we pre-encoded all parameters necessary for sequential, hierarchical assembly (Fig. 1b). The volume ratio of the core-forming segments,  $V_A/V_B$ , determines CBB valence, and the total molecular mass,  $M_n$ , controls particle size<sup>23</sup>. Thus,  $V_A/V_B > 1$  yields monovalent **AB<sup>C</sup>** Janus CBBs with one attractive A patch and one repulsive C patch on opposing sides of the B core (we use bold lettering to distinguish CBBs from the underlying polymer chains with regular lettering);  $V_A/V_B < 1$  leads to divalent **AB<sup>C</sup>A** with two attractive A patches on opposing sides of the B core and a repulsive C patch emanating radially from the B core. The superscript indicates that the C corona is attached to the B core (Extended Data Table 1). From here onwards we replace ABC by the actual block sequences: SBM (polystyrene-*b*-polybutadiene-*b*-poly(methyl methacrylate)) and SDM (polystyrene-*b*-poly(3-butenyl(dodecyl)sulfane)-*b*-poly(methyl methacrylate)). Monovalent **SB<sup>M</sup>** and divalent **SD<sup>M</sup>S** CBBs were prepared separately by self-assembly in *N,N*-dimethylacetamide (DMAc), a non-solvent for the middle blocks, B and D (Extended Data Fig. 1). At this stage the S and M patches are still soluble and not yet completely phase-separated, and they reorganize dynamically into fully developed CBBs when the S block is made insoluble. We trigger this step by changing the solvent to acetone/propan-2-ol (60:40 v/v), leading to particles with essentially different patch arrangements. **SB<sup>M</sup>** and **SD<sup>M</sup>S** thereby act as monomeric units for self-assembly and co-assembly on the next level (transmission electron microscopy (TEM)

images in Fig. 2, level 1, and Extended Data Fig. 2). If kept separate, both units undergo self-assembly into spherical and linear superstructures to minimize energetically unfavourable S patch/non-solvent interfaces (Fig. 2, level 2 self-assembly), once the solvent quality has been reduced to a critical threshold for the S patches. Small corona volumes,  $V_M$ , provide less steric repulsion and promote higher aggregation numbers of **SB<sup>M</sup>** units per spherical (**SB<sup>M</sup>**)<sub>x</sub> cluster or, similarly, higher degrees of polymerization of **SD<sup>M</sup>S** units per [**SD<sup>M</sup>S**]<sub>m</sub> supracolloidal polymer chain (dimer→oligomer→polymer)<sup>23</sup>.

Co-assembly requires at least two CBBs differing in size (10–100 nm), patchiness or chemistry (core/corona). To demonstrate our concept, we mixed two CBBs differing in all three aspects in DMAc in defined particle ratios (**s-SB<sup>M</sup>**:**SD<sup>M</sup>S** = 8:1; Fig. 1 and Extended Data Fig. 2)<sup>9,10</sup> (the prefix 's' in **s-SB<sup>M</sup>** stands for small). We chose small monovalent **s-SB<sup>M</sup>** (hydrodynamic radius  $R_h \approx 10$  nm) in combination with much larger divalent **SD<sup>M</sup>S** ( $R_h \approx 50$  nm). In DMAc, the core-forming segments (B and D) are immiscible; the two CBB species therefore do not exchange terpolymer chains and so evolve independently. As observed for individual species, changes in solvent quality for the S patch destabilize the **SD<sup>M</sup>S** units and induce aggregation into supracolloidal polymer chains with an [**SD<sup>M</sup>S**]<sub>m</sub> sequence. On a similar timescale, the **s-SB<sup>M</sup>** units start to aggregate, yet instead of spherical clusters (self-assemblies), they selectively attach to the newly formed free surface of –S– segments within [**SD<sup>M</sup>S**]<sub>m</sub>, thereby decreasing the –S–/non-solvent interface (Fig. 2, level 2 co-assembly). In the presence of both particles we observe exclusively co-assembly under these conditions. This is surprising, because aggregation of both CBBs is driven solely by a weak non-directional force (solvophobicity) and each CBB is able to form stable populations of spherical and linear superstructures by themselves. Yet the development of [**SD<sup>M</sup>S**]<sub>m</sub> supracolloidal polymer chains favours the attachment of **s-SB<sup>M</sup>** units to the –S– segments. Therefore, we associate this phenomenon with a certain level of cooperativity. We chose **s-SB<sup>M</sup>** units with a particle diameter ( $d_{CBB} \approx 19$  nm) matching the width of the –S– segments of **SD<sup>M</sup>S** ( $w \approx 24$  nm) (Extended Data Fig. 2) and indeed found a defined number of seven to nine **s-SB<sup>M</sup>** units radially covering the –S– segments, which perfectly reflects the original mixing ratio. On exceeding the loading capacity of the –S– segments (for example **s-SB<sup>M</sup>**:**SD<sup>M</sup>S** = 35:1), single **s-SB<sup>M</sup>** CBBs or raspberry-like (**s-SB<sup>M</sup>**)<sub>x</sub> self-assemblies locate in the vicinity of the fully decorated co-assemblies (Extended Data Fig. 3).

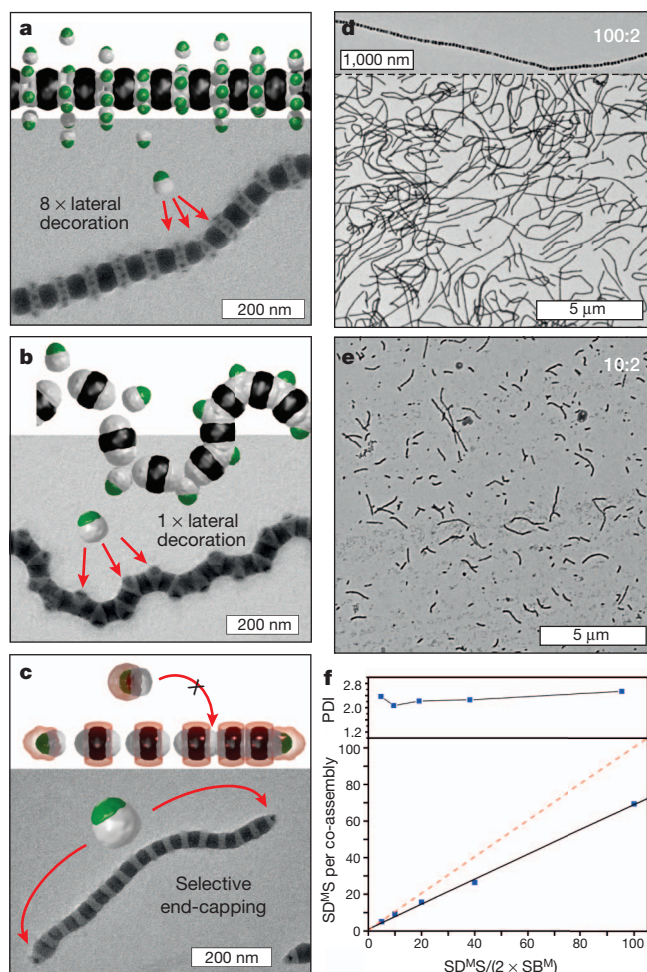
Besides suitable particle ratios, the timescales of aggregation during solvent exchange have to correlate to facilitate proper co-assembly. Because the corona volume of the **s-SB<sup>M</sup>** units affects the critical solvent composition needed for aggregation, we used terpolymers with different lengths of the M block (y stands for s (small), m (medium) or l (large) in Extended Data Table 1 and Extended Data Figs 4 and 5). Divalent **SD<sup>M</sup>S** CBBs self-assemble under solvent conditions in which most **s-SB<sup>M</sup>** prevail as 'monomeric' units. Stability against aggregation is best for **s-SB<sup>M</sup>** with the largest M corona. Whereas **s-SB<sup>M</sup>** leads to simultaneous co-assembly, **s-SB<sup>M</sup>** shows a slight delay. However, in both cases about eight units decorate the –S– segments, pointing to a robust process with sufficient dynamics for rearrangements. However, particles with too short a corona, for example **s-SB<sup>M</sup>**, are unstable before **SD<sup>M</sup>S** polymerization, and lateral decoration is absent. Instead, fully developed spherical (**s-SB<sup>M</sup>**)<sub>x</sub> clusters are incorporated into the linear [**SD<sup>M</sup>S**]<sub>m</sub> superstructure during the step-growth polymerization of **SD<sup>M</sup>S** (Extended Data Fig. 4).

The overall particle size, **x-SB<sup>M</sup>**, of the CBBs (x stands for s (small), m (medium) or l (large)) markedly affects the number of particles attached to each –S– segment of the [**SD<sup>M</sup>S**]<sub>m</sub> supracolloidal polymer chains (Fig. 3 and Extended Data Fig. 6). The particle diameter of **s-SB<sup>M</sup>** units,  $d_{CBB} \approx 19$  nm, allows the incorporation of eight particles on average, also slightly enlarging the width of the –S– segments from  $w \approx 24$  nm to  $w \approx 30$  nm (Figs 2 and 3a). Larger CBBs such as **m-SB<sup>M</sup>** ( $d_{CBB} \approx 35$  nm) require more space; only one particle is therefore able to attach to each –S– segment (Fig. 3b). As a result of their size and the



**Figure 2 | Guided co-assembly across multiple hierarchical levels.** Level 0: triblock terpolymers are the basic building blocks. Level 1: self-assembly of monovalent **s-SB<sup>M</sup>** and divalent **SD<sup>M</sup>S** CBBs in dependence on the volume ratios of the core-forming blocks (intermediate CBBs were captured by crosslinking; Extended Data Fig. 2). Level 2 self-assembly: **s-SB<sup>M</sup>** forms spherical (**SB<sup>M</sup>**)<sub>x</sub> clusters and **SD<sup>M</sup>S** linear [**SD<sup>M</sup>S**]<sub>m</sub> supracolloidal polymer chains. Level 2 co-assembly: **s-SB<sup>M</sup>** and **SD<sup>M</sup>S** with mutually attracting S patches co-assemble into mixed superstructures stabilized by the common M corona. (OsO<sub>4</sub> staining: –S– segments grey, B cores dark grey, –D– segments black, and M not visible as a result of degradation by the electron beam.)





**Figure 3 | Size-selective attachment and control of supracolloidal polymer chain length.** **a**, The  $-S-$  segments accommodate seven to nine  $s-SB^M$  units. **b**, Larger  $m-SB^M$  units fit only once and induce strong kinks. **c**,  $l-SB^M$  units are too large for lateral decoration and instead act as selective end-cappers. **d**, **e**, Length control of supracolloidal polymer chains by the mixing of 100 (**d**) and 10 (**e**)  $SD^M/S$  units per two  $l-SB^M$  end-cappers (Extended Data Fig. 7). Inset shows  $[SD^M/S]_m$  nanostructure. **f**, Polydispersity indices (PDI; top) and linear dependence of  $SD^M/S$  repeating units on the mixing ratio (bottom).

accompanied volume displacement inside the  $-S-$  segments,  $m-SB^M$  units induce strong kinks altering the flexibility and linearity of the co-assembly, which could potentially influence the rheological characteristics and colloidal chain packing.  $l-SB^M$  units substantially exceeding the width of the  $-S-$  segments ( $d_{CBB} \approx 47$  nm) are too large for lateral decoration and specifically locate on terminal positions (Fig. 3c). The relative sizes of the monovalent units and the  $-S-$  segment of the divalent units govern the location and loading capacity, whereas the size of the stabilizing M patch determines the timescales and extent of co-assembly.

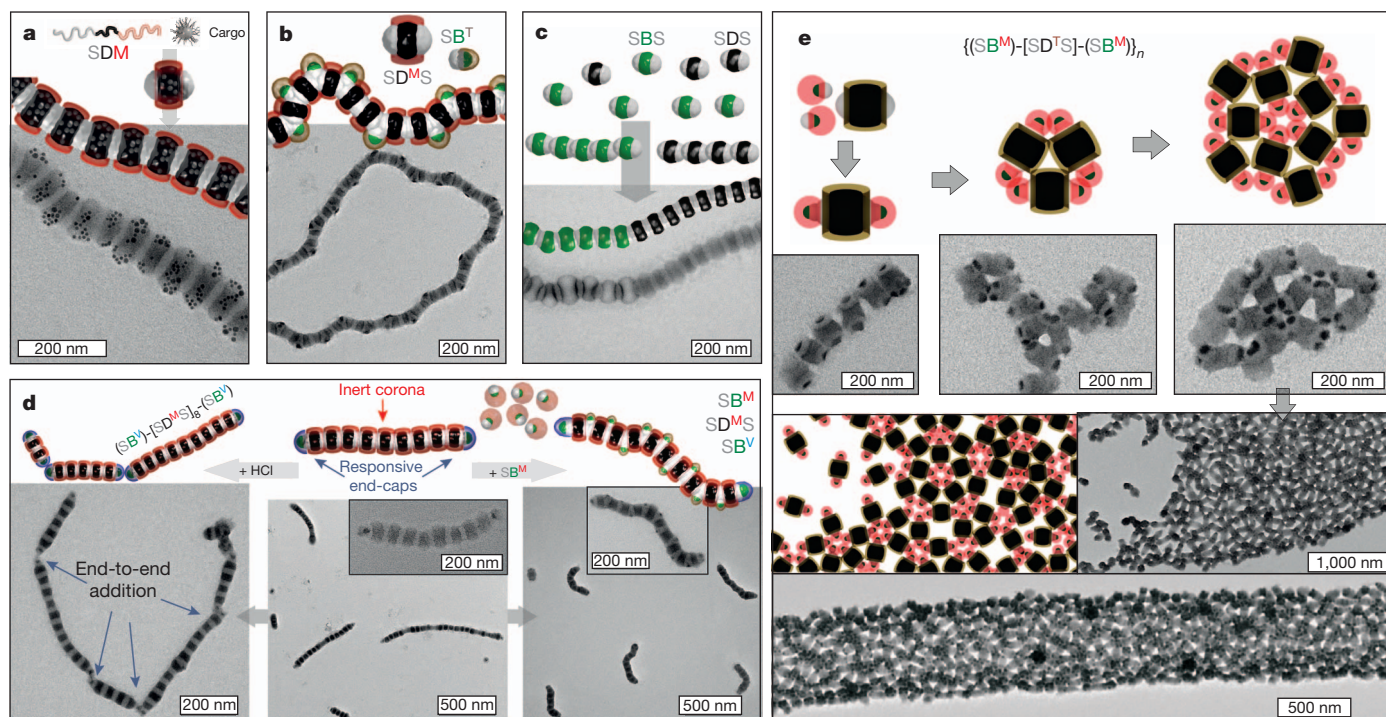
Selective end-capping provides an attractive handle to control the length of the  $[SD^M/S]_m$  supracolloidal polymer chains. Adding  $l-SB^M$  'stoppers' in specific ratios also allows end-functionalization with chemically different CBBs. Without any end-capper, the  $SD^M/S$  units grow into extended structures in the region of  $30 \mu\text{m}$  long (more than 600 repeating units; Extended Data Fig. 7), reaching scales visible by optical microscopy (Extended Data Fig. 6). Figure 3d–f summarizes the dependence of the length distribution of the co-assemblies on the mixing ratio  $SD^M/S:l-SB^M$  (Extended Data Fig. 7). The decrease in mean length is evident when comparing the TEM images in Fig. 3d, e, which show mixing ratios of 100:2 and 10:2. Stoichiometric control as known from step-growth polymerizations is corroborated by the linear dependence

of the average degree of polymerization on the mixing ratio (Fig. 3f). The deviation from the theoretical values (dashed line) is caused by residual 'defects' in the CBBs, because we find a small proportion (less than 1%) of trivalent  $S_3D^M$  CBBs that induce branching. Matching patch sizes of CBBs and low corona volumes of the  $SB^M$  units are decisive in efficient end-capping: the first effectively prevents the addition of  $SD^M/S$  monomeric units and the latter causes the  $SB^M$  units to interfere directly with  $SD^M/S$  polymerization. Both factors can be programmed molecularly into the  $SB^M$  units by variation of the total molecular mass (particle size) and the fraction of M (onset of aggregation).

Controlled mixing of different CBBs permits the rational design and precise implementation of functionalities within the core and/or corona of the co-assemblies (Fig. 4). We prepare hybrid materials by selectively encapsulating 10-nm maghemite nanoparticles within the  $-D-$  segments of  $SD^M/S$  CBBs on hierarchy level 1 (Fig. 4a). Selective and reversible loading yields a unique advantage compared with top-down approaches that require the synthesis of tailored colloids. The preloaded CBBs then polymerize further to linear chains with alternating D-loaded/S-empty compartments. Nanoparticles with tailored affinities for other phases may allow the preparation of bar-coded distributions, and magnetically responsive co-assemblies may serve as advanced viscosity modifiers or may act as deterministic carrier systems. We also generated co-assemblies comprising five different environments (three core/two corona compartments; Fig. 4b) by mixing  $SD^M/S$  with  $SB^T$  ( $T = \text{poly}(t\text{-butyl methacrylate})$ ), which leads to the already established  $[SD^M/S]_m$  core sequence, yet now with a segmented M/T corona. In principle, such structures allow us to control the positioning of nanoparticles along the co-assemblies or to implement predefined responsive folding sites. At this point we emphasize enhanced superstructure stability, because the pristine  $[SD^M/S]_m$  immediately precipitates in ethanol (non-solvent for M) and the  $SB^T/[SD^M/S]_m$  co-assemblies remain stable (Extended Data Fig. 8).

We extended this approach and designed multiblock co-assemblies by combining two divalent units,  $SD^M/S$  and  $SB^M/S$  (Fig. 4c). Here, several short segments of  $[SD^M/S]_m$  or  $[SB^M/S]_n$  sequences are coupled into multiblock co-assemblies. We suggest that the homogeneity and sequence distribution depend on how synchronized self-assembly and co-assembly occur and that deeper kinetic studies will eventually enable control over the sequence length. In an analogy with polymer architectures, we fabricated telechelic oligomers with terminal functional groups by end-capping  $[SD^M/S]_{10}$  oligomers with  $SB^V$  ( $V = \text{poly}(2\text{-vinylpyridine})$ ; Fig. 4d). The number-average length of the telechelics scales with the mixing ratio  $SD^M/S:SB^V$ . The V corona is clearly visible as grey end-cap by TEM (Fig. 4d, central inset). Protonation of the V corona with HCl causes solvophobic attraction and triggers aggregation of the V-termini into extended subdivided superstructures. We emphasize that this process is fully reversible and that it corresponds to a step-growth polymerization on two levels: first for the individual  $SD^M/S$  units, and then chain extension of the telomerized 'macromonomers'. Combining selective end-capping and lateral decoration yields ternary structures. The addition of  $SB^M$  units to the  $SB^V-[SD^M/S]_{10}-SB^V$  telechelic oligomers does indeed lead to the decoration of the  $-S-$  segments and a final composition of  $SB^M:SD^M/S:SB^V = 10:10:2$ .

Finally, we bridge three hierarchical levels by co-assembly of  $SB^M$  and  $SD^T/S$  building blocks into end-capped colloidal molecules with the composition  $(SB^M)_N-[SD^T/S]_1-(SB^M)_N$  (where the number of end-caps  $N = 1-5$ ). The CBB mixing ratio determines the number of end-cappers; for example,  $SB^M:SD^T/S = 2:1$  gives  $(SB^M)_1-[SD^T/S]_1-(SB^M)_1$ . These co-assemblies self-assemble spontaneously into triangles and further into two-dimensional networks when cast onto substrates, a necessary two-dimensional confinement tool<sup>13,30</sup>, and the number of end-caps ( $SB^M$ ) directs the number of nearest neighbours at the network linkages (Fig. 4e). The self-assembly of these colloidal molecules is directed by the M/T corona patches that change their volume on solvent evaporation and develop into attractive M and repelling T patches. For  $N = 1$  or 2 the end-cap size allows three or four nearest



**Figure 4 | Binary and ternary co-assemblies.** **a**, Hybrid co-assemblies with magnetite nanoparticles selectively loaded into one core segment. **b**, Core and corona co-assemblies. The  $[\text{SDM}]_m$  superstructure is decorated with  $\text{SB}^T$  CBBs, yielding an alternating M/T corona. **c**, Linear multiblock co-assembly composed of  $[\text{SDM}]_m/[\text{SBM}]_n$  sequences (M corona omitted for clarity). **d**,  $\text{SB}^V$  end-cappers produce telechelic  $[\text{SDM}]_{10}$  oligomers (grey V

corona surrounding the  $\text{SB}^V$  end-cap in inset). Protonation with HCl triggers further polymerization, and decoration with  $\text{SB}^M$  units yields ternary co-assemblies. **e**, Multicomponent structuring across three hierarchical levels. In solution, the terpolymers self-assemble into CBBs that co-assemble into colloidal molecules; these finally self-assemble into networks after drop-casting on a substrate.

neighbours,  $\{\text{SB}^M\text{-SD}^T\text{S-}\text{SB}^M\}_{4,5}$ , whereas for  $N=3-5$  there is only space for two nearest neighbours and the networks show the onset of distorted kagome lattice formation (Extended Data Fig. 9).

These hierarchical multicomponent superstructures demonstrate the level of complexity reached through the controlled co-assembly of soft patchy nanoparticles and is in stark contrast to the self-assembly of increasingly complex building blocks (shape and surface pattern). Using proper interacting segments, a wide range of building block combinations is conceivable, from biological origin to inorganic and organic nanoparticles and microparticles. Various self-assembly stimuli can conveniently be implemented by means of functional polymer blocks, for example supramolecular interactions, chemical reactions and environmental triggers (solvent polarity, pH, temperature, light or electrochemistry). Co-assembly opens up avenues to construct new materials, also far from thermodynamic equilibrium, through the tailored spatial organization of functionalities and the control of kinetics. We foresee application possibilities in smart materials, sensing, photonics and nanolithography.

## METHODS SUMMARY

Relevant parameters of SBM, SBV and SBT triblock terpolymers are summarized in Fig. 1 and Extended Data Table 1. SBM and SBT were converted to SDM and SDT by thiol-ene click reaction of 1-dodecane thiol to the poly(1,2-butadiene) block. All CBBs were prepared in DMAc at a polymer concentration of  $0.1 \text{ g l}^{-1}$  and annealed overnight at  $70^\circ\text{C}$ . Particle dispersions were mixed in specific particle ratios and 10 ml of solution was co-dialysed (molecular mass cut-off 12,000–14,000  $\text{g mol}^{-1}$ ; Roth) against 5 l of selective solvent or solvent mixture for the corona block (acetone/propan-2-ol for M and V; ethanol for T corona). The solvent exchange was monitored by  $^1\text{H-NMR}$ . TEM was performed in bright-field mode on Zeiss CEM 902 and 922 OMEGA electron microscopes operated at 80 kV and 200 kV, respectively. Co-assemblies were deposited by drop-casting (0.05 ml of  $0.1 \text{ mg ml}^{-1}$  dispersion) onto carbon-coated copper grids resting on a filter paper to blot excess solution immediately. Samples were exposed to  $\text{OsO}_4$  vapour for 2 h to stain the polybutadiene blocks selectively.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 March; accepted 28 August 2013.

Published online 3 November 2013.

- Whitesides, G. M. & Grzybowski, B. Self-assembly at all scales. *Science* **295**, 2418–2421 (2002).
- Stuart, A. R. Towards high-performance bioinspired composites. *Adv. Mater.* **24**, 5024–5044 (2012).
- Fratzl, P. & Weinkammer, R. Nature's hierarchical materials. *Prog. Polym. Sci.* **52**, 1263–1334 (2007).
- Mann, S. Self-assembly and transformation of hybrid nano-objects and nanostructures under equilibrium and non-equilibrium conditions. *Nature Mater.* **8**, 781–792 (2009).
- Glötzer, S. C. & Solomon, M. J. Anisotropy of building blocks and their assembly into complex structures. *Nature Mater.* **6**, 557–562 (2007).
- Quan, Z. & Fang, J. Superlattices with non-spherical building blocks. *Nano Today* **5**, 390–411 (2010).
- Wang, T. *et al.* Self-assembled colloidal superparticles from nanorods. *Science* **338**, 358–363 (2012).
- Grzelczak, M., Vermant, J., Furst, E. M. & Liz-Marzán, L. M. Directed self-assembly of nanoparticles. *ACS Nano* **4**, 3591–3605 (2010).
- Nie, Z. *et al.* Self-assembly of metal-polymer analogues of amphiphilic triblock copolymers. *Nature Mater.* **6**, 609–614 (2007).
- Liu, K. *et al.* Step-growth polymerization of inorganic nanoparticles. *Science* **329**, 197–200 (2010).
- Cui, H., Chen, Z., Zhong, S., Wooley, K. L. & Pochan, D. J. Block copolymer assembly by kinetic control. *Science* **317**, 647–650 (2007).
- Chen, Q. *et al.* Supracolloidal reaction kinetics of Janus spheres. *Science* **331**, 199–202 (2011).
- Chen, Q., Bae, S. C. & Granick, S. Directed self-assembly of a colloidal kagome lattice. *Nature* **469**, 381–384 (2011).
- Wang, Y. *et al.* Colloids with valence and specific directional bonding. *Nature* **491**, 51–55 (2012).
- Kaufmann, T. *et al.* 'Sandwich' microcontact printing as a mild route towards monodisperse Janus particles with tailored bifunctionality. *Adv. Mater.* **23**, 79–83 (2010).
- Chen, Q., Bae, S. C. & Granick, S. Staged self-assembly of colloidal metastructures. *J. Am. Chem. Soc.* **134**, 11080–11083 (2012).
- Li, Z., Kesselman, E., Talmon, Y., Hillmyer, M. A. & Lodge, T. P. Multicompartiment micelles from ABC miktoarm stars in water. *Science* **306**, 98–101 (2004).



18. Wang, X. *et al.* Cylindrical block copolymer micelles and co-micelles of controlled length and architecture. *Science* **317**, 644–647 (2007).
19. Schmelz, J., Schedl, A. E., Steinlein, C., Manners, I. & Schmalz, H. Length control and block-type architectures in worm-like micelles with polyethylene cores. *J. Am. Chem. Soc.* **134**, 14217–14225 (2012).
20. Gädt, T., Ieong, N. S., Cambridge, G., Winnik, M. A. & Manners, I. Complex and hierarchical micelle architectures from diblock copolymers using living, crystallization-driven polymerizations. *Nature Mater.* **8**, 144–150 (2009).
21. Kubowicz, S. *et al.* Multicompartment micelles formed by self-assembly of linear ABC triblock copolymers in aqueous medium. *Angew. Chem. Int. Ed.* **44**, 5262–5265 (2005).
22. Gröschel, A. H. *et al.* Facile, solution-based synthesis of soft, nanoscale Janus particles with tunable Janus balance. *J. Am. Chem. Soc.* **134**, 13850–13860 (2012).
23. Gröschel, A. H. *et al.* Precise hierarchical self-assembly of multicompartment micelles. *Nature Commun.* **3**, 710, <http://dx.doi.org/10.1038/ncomms1707> (2012).
24. Li, Z., Hillmyer, M. A. & Lodge, T. P. Control of structure in multicompartment micelles by blending  $\mu$ -ABC star terpolymers with AB diblock copolymers. *Macromolecules* **39**, 765–771 (2005).
25. Fang, B. *et al.* Undulated multicompartment cylinders by the controlled and directed stacking of polymer micelles with a compartmentalized corona. *Angew. Chem. Int. Ed.* **48**, 2877–2880 (2009).
26. Ruper, P. A., Chabanne, L., Winnik, M. A. & Manners, I. Non-centrosymmetric cylindrical micelles by unidirectional growth. *Science* **337**, 559–562 (2012).
27. Chen, Q., Yan, J., Zhang, J., Bae, S. C. & Granick, S. Janus and multiblock colloidal particles. *Langmuir* **28**, 13555–13561 (2012).
28. Sacanna, S., Irvine, W. T. M., Chaikin, P. M. & Pine, D. J. Lock and key colloids. *Nature* **464**, 575–578 (2010).
29. Bates, F. S. *et al.* Multiblock polymers: panacea or Pandora's box? *Science* **336**, 434–440 (2012).
30. Bowden, N., Terfort, A., Carbeck, J. & Whitesides, G. M. Self-assembly of mesoscale objects into ordered two-dimensional arrays. *Science* **276**, 233–235 (1997).

**Acknowledgements** We thank O. Ikkala and E. Kumacheva for discussions and comments on the manuscript, and A. Majewski for providing maghemite nanoparticles. This work was supported by Deutsche Forschungsgemeinschaft within Sonderforschungsbereich 840 (TP A1 and A2).

**Author Contributions** A.H.G. initiated the project. A.H.G. and T.I.L. performed experiments and collected data. A.H.G., A.W. and A.H.E.M. designed the experiments, discussed results and wrote the manuscript. F.H.S. co-designed experiments, discussed results and commented on the manuscript. H.S. provided polymers, discussed results and commented on the manuscript. A.H.E.M. supervised the project.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.H.E.M. ([axel.mueller@uni-mainz.de](mailto:axel.mueller@uni-mainz.de)) or A.H.G. ([andre.groschel@aalto.fi](mailto:andre.groschel@aalto.fi)).



## METHODS

**Particle preparation.** All terpolymers were synthesized by sequential anionic polymerization<sup>31–35</sup>. Important parameters are summarized in Fig. 1 and Extended Data Tables 1. All solvents used were of analytical grade. Dialysis tubes of regenerated cellulose with a molecular mass cut-off of 12,000–14,000 g mol<sup>−1</sup> were purchased from Roth. Oleic acid-stabilized maghemite nanoparticles were synthesized as described elsewhere<sup>36</sup>. Both polymers, SBM and SBT, were converted to SDM and SDT, respectively, by means of a thiol-ene click reaction of 1-dodecane thiol with the pendant double bonds of poly(1,2-butadiene). In a typical experiment, terpolymer (0.5 g) was dissolved in 20 ml of tetrahydrofuran; 20 ml of 1-dodecane thiol were added and the solution was purged for 15 min with argon. The mixture was irradiated for 24 h with an ultraviolet lamp with a cut-off filter ( $\lambda = 300$  nm;  $\lambda_{\text{max}} = 360$  nm). After the reaction, the modified terpolymer was precipitated in propan-2-ol, filtered, and washed with excess propan-2-ol to remove remaining 1-dodecane thiol. SBM, SDM, SBT, SDT and SBV triblock terpolymers were dissolved separately in DMAc at an initial polymer concentration of 0.1 g l<sup>−1</sup> and annealed overnight at 70 °C to guarantee an equilibrated system. At this stage of self-assembly the corona patches were not yet fully developed. To reveal the patches in the developed state (Extended Data Fig. 2), the CBBs were crosslinked while located within the respective superstructure. Therefore, 2 equivalents of the photo-crosslinker 2,4,6-trimethylbenzoyldiphenylphosphineoxide (Lucirin TPO;  $\lambda_{\text{max}} \approx 360$  nm), were added to each double bond. Gentle stirring for 2 h ensured a homogeneous distribution of the photo-crosslinker before the samples were irradiated for 1 h with an ultraviolet lamp with a cut-off filter ( $\lambda = 300$  nm). Redispersion in DMAc as a good solvent for the patches then broke up the superstructures into the respective CBBs.

**Self-assembly and co-assembly.** CBB solutions (0.1 g l<sup>−1</sup> in DMAc) were mixed in specific particle ratios to yield 10 ml of a colloidal mixture and co-dialysed against 5 l of selective solvent/solvent mixture for the corona block (acetone/propan-2-ol for the M corona and V corona block, and ethanol for the T-corona block). The solvent exchange was monitored by <sup>1</sup>H-NMR. TEM was performed in bright-field mode on Zeiss CEM 902 and LEO 922 OMEGA electron microscopes operated at 80 kV and 200 kV, respectively. Samples were prepared by dropping 0.05 ml of a 0.1 mg ml<sup>−1</sup> colloidal suspension onto carbon coated copper grids resting on a filter paper to remove the excess solution immediately. The two-dimensional networks were prepared similarly, except from 0.01 g l<sup>−1</sup> colloidal suspension that was allowed to settle on the TEM grid for 30 s before blotting. TEM grids were then exposed for 2 h to OsO<sub>4</sub> vapour to stain the polybutadiene block selectively. The number of polymer chains in each CBB was determined, evaluating at least 250 cores. The average degree of polymerization of each [SD<sup>M</sup>S]<sub>m</sub> colloidal polymer was determined by counting the black segments of 500 colloidal polymers and plotted as frequency distribution.

**Correlating co-assembly with onsets of aggregation.** We used a CBB ratio of s-SB<sup>M(y)</sup>.SD<sup>M</sup>S = 8:1 for all experiments to ensure that enough monovalent CBBs were present for full decoration of the –S– segments. SD<sup>M</sup>S was co-dialysed with three different s-SB<sup>M(y)</sup> units with varying volume of the corona, M(y) (y stands for s (small), m (medium) or l (large)). The dialysis sequence was from DMAc into acetone/propan-2-ol mixtures. Greater amounts of propan-2-ol led to a contraction of the M corona. SD<sup>M</sup>S forms extended superstructures irrespective of the propan-2-ol content (also in pure acetone). s-SB<sup>M(l)</sup> CBBs with a large relative volume fraction of the soluble M corona of  $r_M = 0.74$  (where  $r_M = V_M/(V_S + V_B)$ ) do not co-assemble at low propan-2-ol contents (less than 20 vol%) and are only located in the vicinity of polymerized SD<sup>M</sup>S (Extended Data Fig. 4). The long corona blocks provide steric stabilization by fully covering the CBB. At high

propan-2-ol contents (30 vol%), partial co-assembly is observed and quantitative co-assembly only at 40 vol% as a result of progressive contraction of the M corona. Because the SD<sup>M</sup>S units are already fully polymerized, decoration does not interfere with the growth process (no premature end-capping). s-SB<sup>M(m)</sup> CBBs with  $r_M = 0.34$  carry a corona block of moderate length and show partial to quantitative co-assembly over the whole range of solvent mixtures, indicating only a slight difference in onset of aggregation between SD<sup>M</sup>S and s-SB<sup>M(m)</sup>. s-SB<sup>M(s)</sup> CBBs with the shortest corona block,  $r_M = 0.15$ , show no co-assembly in either of the solvent compositions: at a propan-2-ol content of only 10 vol% the corona shows insufficient stabilization for the s-SB<sup>M(s)</sup> CBBs, and aggregation into spherical (s-SB<sup>M(s)</sup>)<sub>x</sub> clusters occurs before polymerization of SD<sup>M</sup>S. The SD<sup>M</sup>S units still polymerize through these spherical clusters, resulting in randomly distributed bulbs of (s-SB<sup>M(s)</sup>)<sub>x</sub> clusters along the [SD<sup>M</sup>S]<sub>n</sub> chain.

**Calculation of polymer chains in each particle and compartment volumes.** Because CBBs form dynamically during the self-assembly of triblock terpolymers, one needs to know how many polymer chains are located within one CBB to be able to mix CBBs in the desired ratios for the co-assembly process. We determined the average core diameter of each CBB by TEM (averaged over 250 samples) to calculate the aggregation number of triblock terpolymer chains,  $N_{\text{agg}}$  in each CBB. This evaluation is exemplified on monovalent spherical s-SB<sup>M</sup> and divalent, cylindrical SD<sup>M</sup>S cores (Extended Data Fig. 2).  $N_{\text{agg}}$  was then calculated from

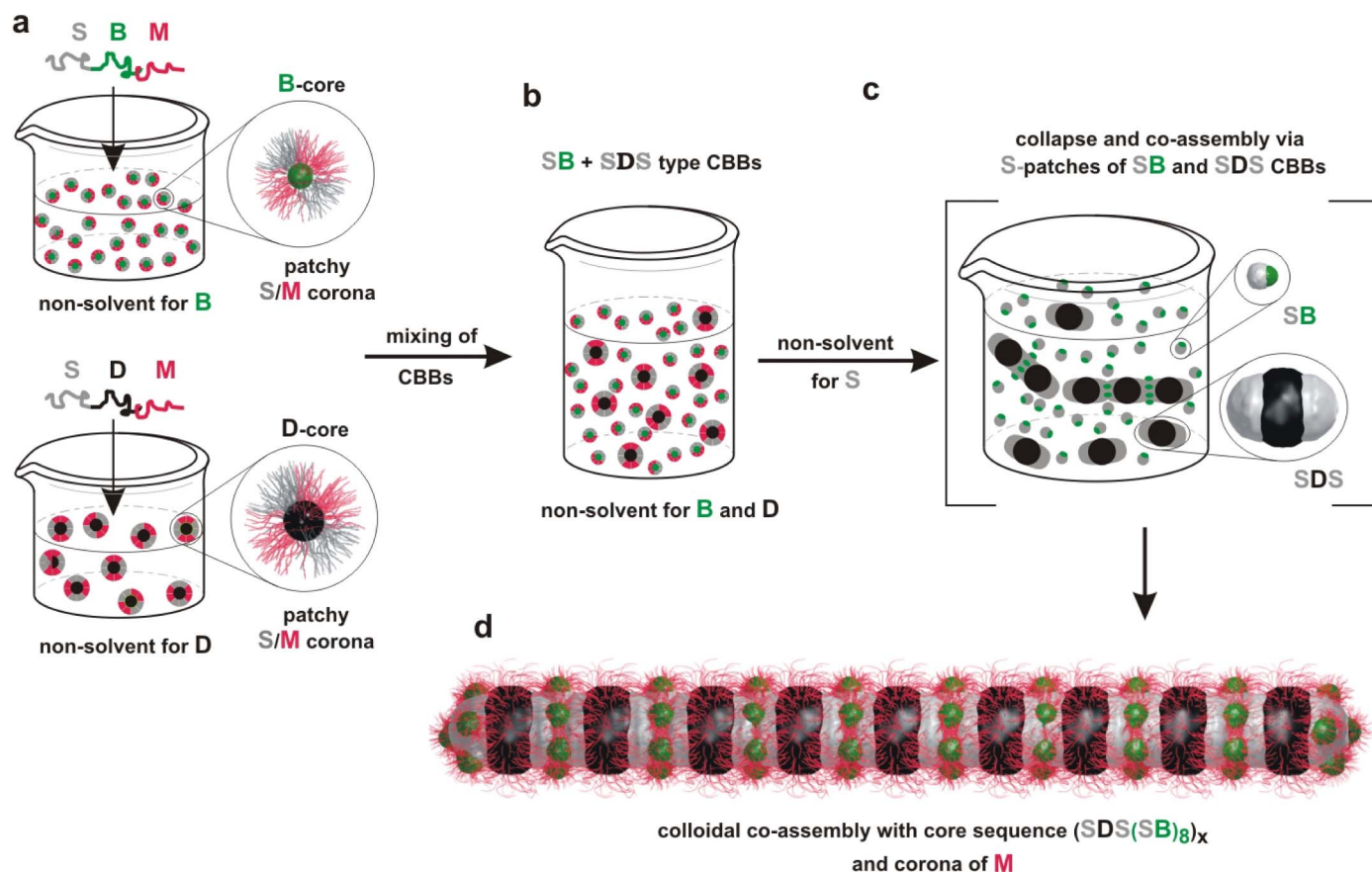
$$N_{\text{agg}} = \frac{m_{\text{core}}}{m_{\text{chain}}} = V_{\text{core}} \frac{N_A \rho_B}{M_{\text{B}}^{\text{chain}}}$$

with  $V_{\text{core}} = 4/3(\pi R_{\text{core}}^3)$  for spherical SB<sup>M</sup> and  $V_{\text{core}} = \pi R_{\text{core}}^2$  for cylindrical SD<sup>M</sup>S;  $m_{\text{core}}$  is the mass of the micellar core,  $m_{\text{B}}^{\text{chain}}$  is the mass of the middle block,  $M_{\text{B}}^{\text{chain}}$  is the molecular mass of the middle block,  $N_A$  is Avogadro's constant and  $\rho_B$  is the density of the middle block. We found that this approach yields a good correlation between the ratio of triblock terpolymers and the final composition of CBBs in the co-assemblies. We can calculate the diameter of SB<sup>M</sup> units combining the calculated volume for the B compartments,  $V_B$ , from  $N_{\text{agg}}$  with the volume ratio  $V_S/V_B$ . The diameter of spherical CBBs,  $d_{\text{CBB}}$ , is then given by

$$d_{\text{CBB}} = 2 \left( \left( V_S + \frac{1}{2} V_B \right) \frac{3}{4\pi} \right)^{\frac{1}{3}}$$

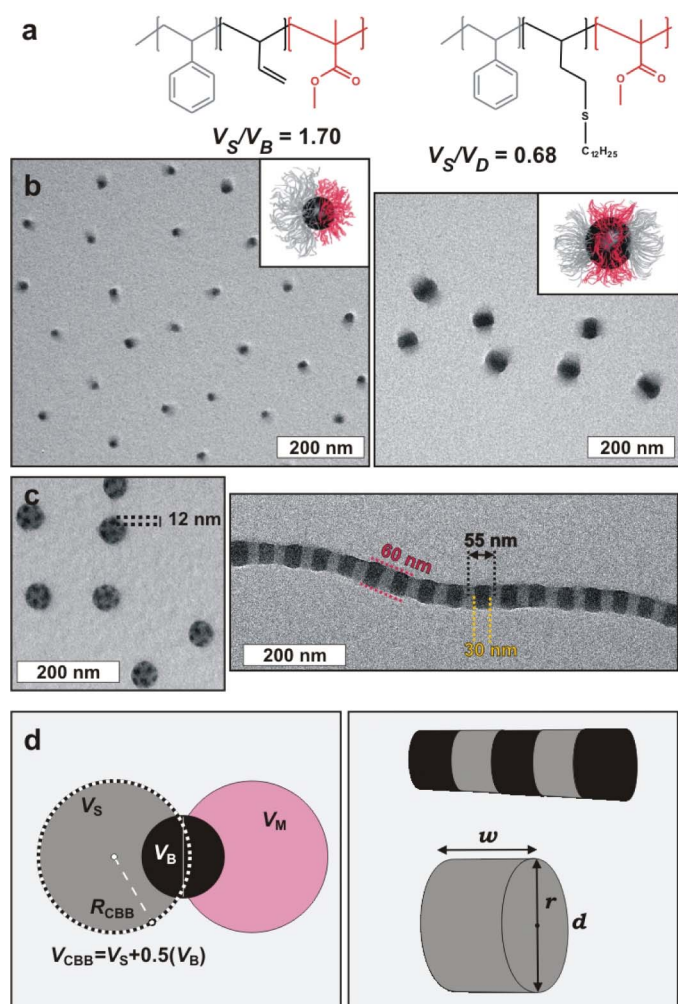
The volume and surface area of the cylindrical –S– segments are simply determined from TEM data; that is, directly from the colloidal co-assembly as width ( $w$ ) and height = diameter ( $d$ ).

1. Auschra, C. & Stadler, R. Synthesis of block copolymers with poly(methyl methacrylate): P(B-*b*-MMA), P(EB-*b*-MMA), P(S-*b*-B-*b*-MMA) and P(S-*b*-EB-*b*-MMA). *Polym. Bull.* **30**, 257–264 (1993).
2. Giebler, E. & Stadler, R. ABC triblock polyampholytes containing a neutral hydrophobic block, a polyacid and a polybase. *Macromol. Chem. Phys.* **198**, 3815–3825 (1997).
3. Walther, A., André, X., Drechsler, M., Abetz, V. & Müller, A. H. E. Janus discs. *J. Am. Chem. Soc.* **129**, 6187–6198 (2007).
4. Schacher, F., Yuan, J., Schoberth, H. G. & Müller, A. H. E. Synthesis, characterization, and bulk crosslinking of polybutadiene-*block*-poly(2-vinyl pyridine)-*block*-poly(*tert*-butyl methacrylate) block terpolymers. *Polymer* **51**, 2021–2032 (2010).
5. Ruckdäschel, H. et al. Compatibilisation of PPE/SAN blends by triblock terpolymers: correlation between block terpolymer composition, morphology and properties. *Polymer* **47**, 2772–2790 (2006).
6. Majewski, A. P. et al. Dual-responsive magnetic core–shell nanoparticles for nonviral gene delivery and cell separation. *Biomacromolecules* **13**, 857–866 (2012).



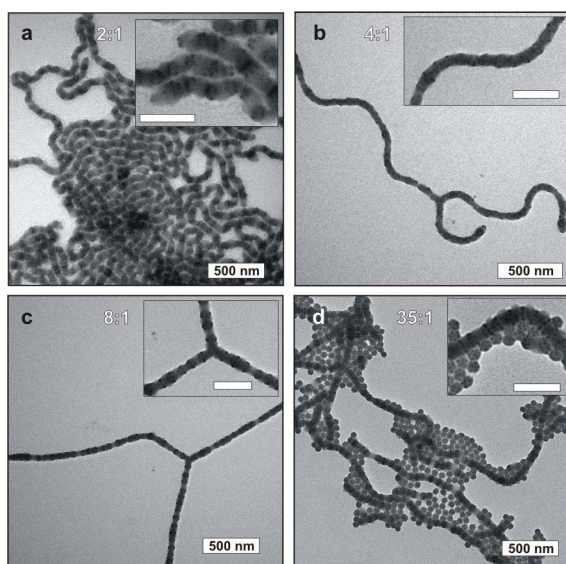
**Extended Data Figure 1 | Experimental approach to prepare colloidal co-assemblies exemplified on s-SB<sup>M</sup> and SD<sup>M</sup>S.** **a**, First, the triblock terpolymers are dispersed separately in DMAc as a non-solvent for B and D, equalling bottom-up structuring of CBBs with different cores (B and D), yet identical ‘sticky’ S patch and stabilizing M corona. **b**, Both colloidal dispersions are

mixed in specific ratios and stirred overnight to ensure homogeneous distribution. **c**, Co-dialysis into a non-solvent for the S patches induces collapse and aggregation. SD<sup>M</sup>S grows into extended linear colloidal polymers decorated by s-SB<sup>M</sup> units (M corona omitted for clarity). **d**, Final colloidal co-assemblies stabilized by the common M corona.

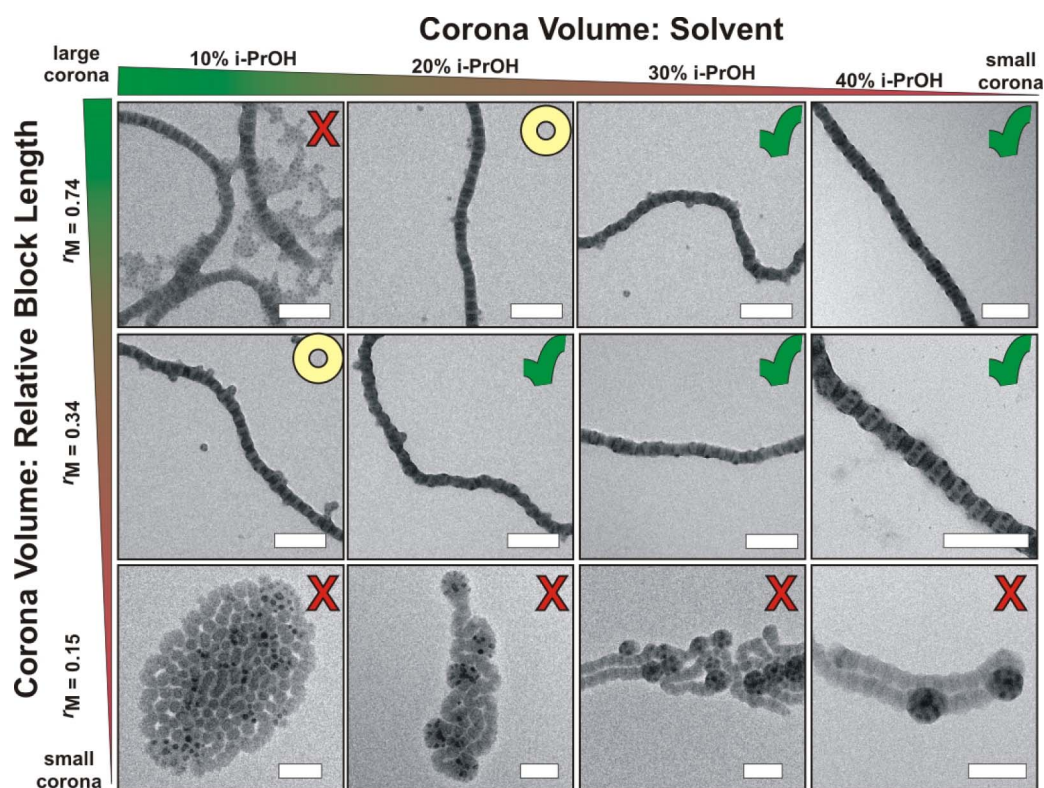


**Extended Data Figure 2 | Compartment sizes, polymer chains per CBB and CBB mixing ratios.** **a**, SBM with  $V_S/V_B = 1.70$  and SDM with  $V_S/V_D = 0.68$  result in **b**, monovalent  $SB^M$  (Janus) and divalent  $SD^M S$  units, respectively. CBBs are visualized after crosslinking of the remaining double bonds within the particle cores. **c**, Examples of spherical  $SB^M$  clusters and linear  $SD^M S$  colloidal polymers. **d**, Diagrams of CBB compartment volumes and surface area of associative patches. We determine the number of  $SB^M$  units able to attach to the  $-S-$  segments of the  $SD^M S$  colloidal polymers by calculating the volume and the diameter of the CBBs, assuming a spherical shape of B and S phases and considering the number of polymer chains per patch. The spherical  $SB^M$  units are composed of (i) the body as the sum of the collapsed S patch (grey) and B core (black) and (ii) the M corona patch. The radius,  $R_{CBB}$ , and the volume,  $V_{CBB}$ , displaced by the CBB when aggregating into the  $-S-$  segments are estimated by combining the volume of S plus half of the volume of B (the dashed line marks  $R_{CBB}$ ). These assumptions are valid for the collapsed state, given the unfavourable interactions with the surrounding medium and the minimization of the interfacial energies.



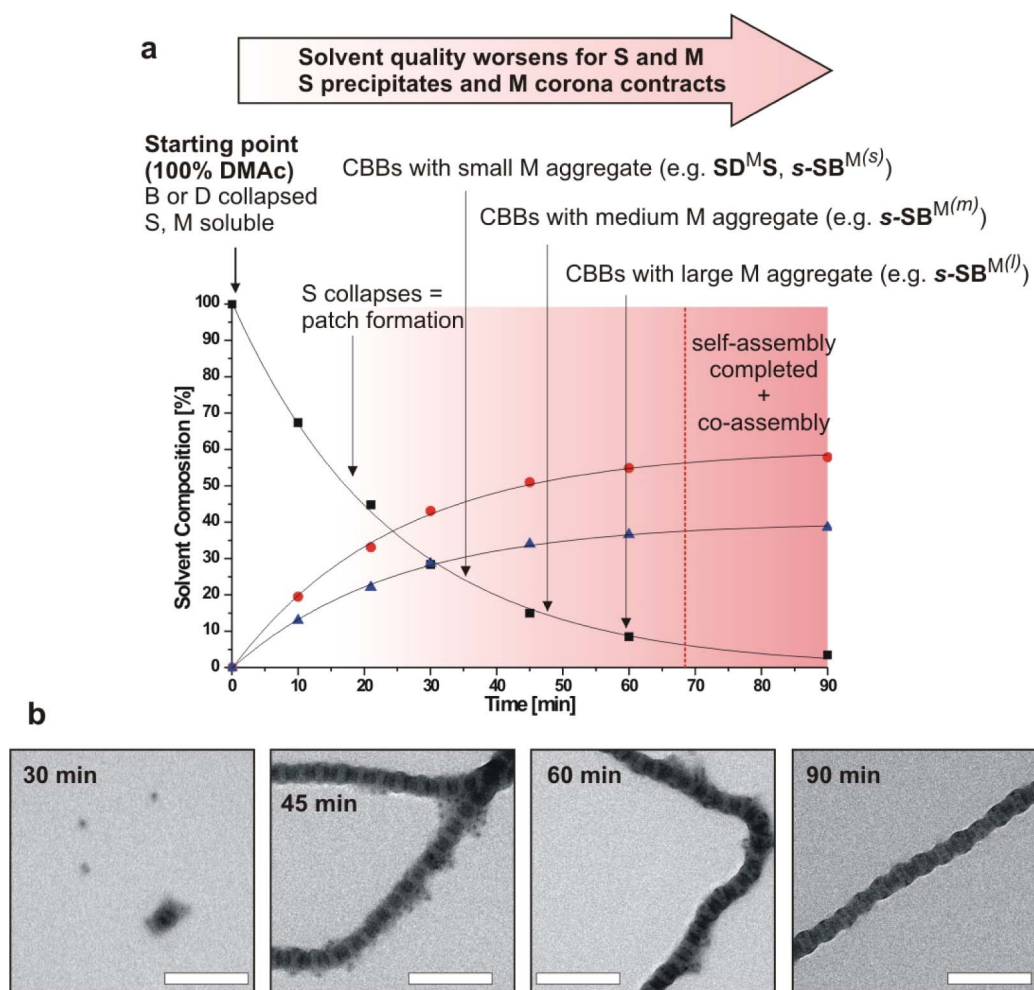


**Extended Data Figure 3 | Loading capacity for decoration with spherical CBBs during co-assembly.**  $\text{SD}^{\text{M}}\text{S}$  and  $\text{s-SB}^{\text{M}}$  colloidal co-assemblies with increasing mixing ratios 2:1, 4:1, 8:1 and 35:1 of  $\text{s-SB}^{\text{M}}:\text{SD}^{\text{M}}\text{S}$  and corresponding co-assembly composition. **a**, At small excess,  $\text{s-SB}^{\text{M}}:\text{SD}^{\text{M}}\text{S} = 2:1$ , the  $-\text{S}-$  segments are only partly decorated. **b**, **c**, With increasing CBB ratio, the  $-\text{S}-$  segments of the colloidal polymers are more strongly occupied (4:1) (**b**) and at full capacity in a radial manner (8:1) (**c**). **d**, Above this ratio, no space remains on the colloidal polymer and only then does  $\text{s-SB}^{\text{M}}$  start to form homo-clusters (35:1; football-like or raspberry-like homo-clusters adjacent to the colloidal co-assemblies). These observations fit well with our calculations, as a cylinder with  $d_{\text{S-}} \approx 55$  nm and  $w \approx 24$  nm can accommodate up to nine  $\text{s-SB}^{\text{M}}$  units with  $d_{\text{CBB}} \approx 19$  nm ( $\pi d_{\text{S-}} \approx 173$  nm;  $173 \text{ nm}/19 \text{ nm} \approx 9.1$ ). Scale bars in insets are 200 nm (OsO<sub>4</sub> staining: S grey, B dark grey dots, D black, and M not visible as a result of degradation by the electron beam.)



**Extended Data Figure 4 | Co-assembly of  $s\text{-SB}^{M(y)}$  and  $\text{SD}^M\text{S}$  in dependence on the corona volume of the monovalent  $s\text{-SB}^{M(y)}$  units ( $y$  stands for  $s$ ,  $m$  or  $l$ ).** Colloidal co-assembly occurs preferentially with matching onsets of aggregation that depend on the corona volume, which is tunable either by decreasing the block length of the corona block,  $r_M = V_M/(V_S + V_B)$ ,  $r_M(s\text{-SB}^{M(l)}) = 0.74$ ,  $r_M(s\text{-SB}^{M(m)}) = 0.34$ ,  $r_M(s\text{-SB}^{M(s)}) = 0.15$ , or by reducing the solvent quality,

here the addition of propan-2-ol from 10 vol% to 40 vol%. Thereby, crosses indicate no co-assembly, circles the onset of co-assembly and tick marks effective and quantitative co-assembly. For detailed discussion see Methods. (OsO<sub>4</sub> staining: S grey, B dark grey dots, D black, and M not visible as a result of degradation by the electron beam.)

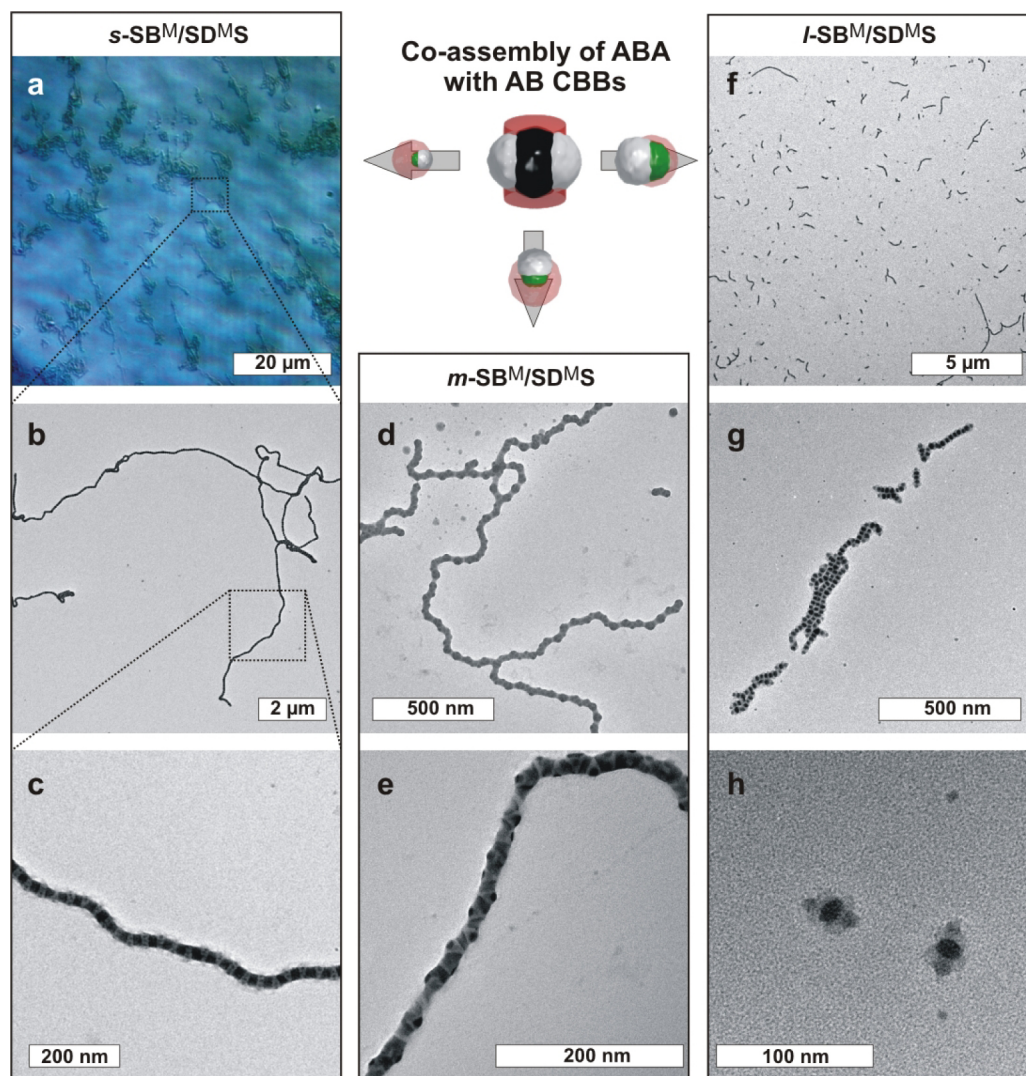


**Extended Data Figure 5 | Timeline of CBB aggregation and response of CBB constituents on solvent composition.** **a**, Co-dialysis of CBBs (here  $SD^M S$  and  $s-SB^{M(y)}$ ) from DMAc (squares) into acetone (circles)/propan-2-ol (triangles) mixtures (60:40 v/v) results in a continuous change of a multitude of polymer–polymer interactions as well as polymer block responses to the ternary solvent mixture. With progressing dialysis, DMAc is replaced by acetone/propan-2-ol, affecting the solubility of S/M corona patches. This induces the collapse into S patches and determines the onset of aggregation. Time-dependent  $^1H$ -NMR measurements were performed by drawing samples

at specific intervals during dialysis to determine the solvent composition.

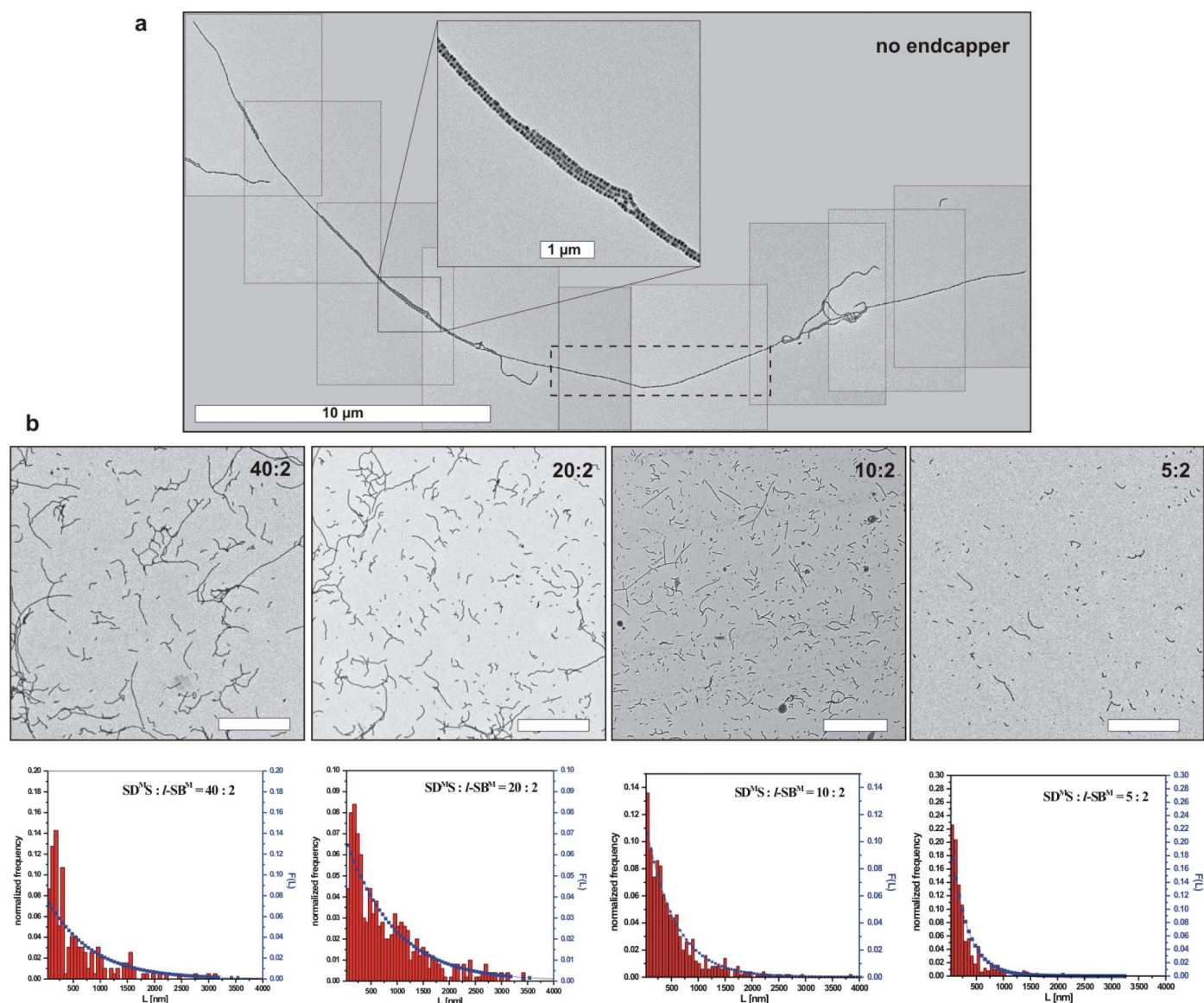
**b**, TEM series exemplifying the timeline of aggregation on  $s-SB^{M(l)}$  and  $SD^M S$ . Up to 30 min, S corona collapses to form S patches, yet both CBBs are still stable due to stabilizing corona; after 30 min,  $SD^M S$  (smaller M corona) aggregate; after 45 min, aggregation of  $SD^M S$  proceeds, while  $s-SB^{M(l)}$  is still stable (larger M corona, no co-assembly); at 60 min,  $s-SB^{M(l)}$  aggregates and co-assembly takes place; up to 90 min, co-assembly is complete. ( $OsO_4$  staining: S grey, B dark grey dots, D black, and M not visible as a result of degradation by the electron beam.)





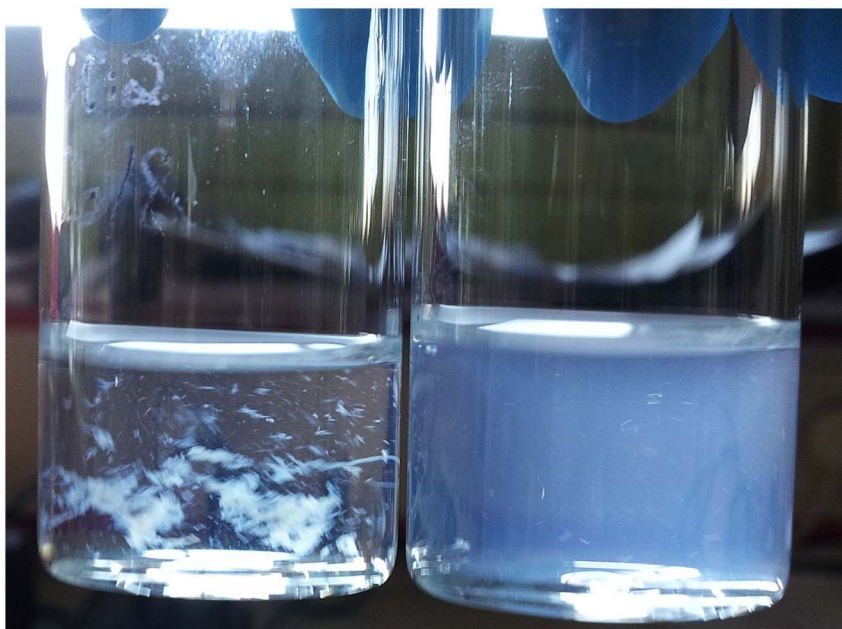
**Extended Data Figure 6 | Supporting images of colloidal co-assemblies composed of  $SD^M S$  and  $s-SB^M$ ,  $m-SB^M$  and  $l-SB^M$ .** **a**, Optical microscopy image of colloidal co-assemblies 30  $\mu m$  in length. **b**, **c**, TEM images of fully decorated  $-S-$  segments with  $s-SB^M$ . **d**, **e**, Larger  $m-SB^M$  units induce kinks due to lateral, alternating decoration and high volume displacement within the

$-S-$  segments. **f**, Zoom-out of rod-like colloidal oligomers end-capped with  $l-SB^M$ . **g**, Magnification of **f**. **h**, The smallest possible co-assembly with  $[SD^M S]_1$ ; that is, two end-caps attached to one  $SD^M S$  unit. (OsO<sub>4</sub> staining: S grey, B dark grey dots, D black, and M not visible as a result of degradation by the electron beam.)



**Extended Data Figure 7 | Length control of colloidal polymers via end-capping with l-SB<sup>M</sup>.** **a**, Without any l-SB<sup>M</sup> end-capper added, the SD<sup>MS</sup> units grow into remarkably extended superstructures several micrometres in length, exceeding 500–600 SD<sup>MS</sup> repeating units. The inset illustrates the segmented core and some occurrence of branching. The displayed image consists of an overlay of nine separate TEM images, because the superstructure was far too large for magnifications that were still able to resolve the nanostructure. The dashed box marks the part of the colloidal polymer that is shown in Fig. 3d. **b**, TEM images and frequency distributions (red bars) of 500 evaluated colloidal

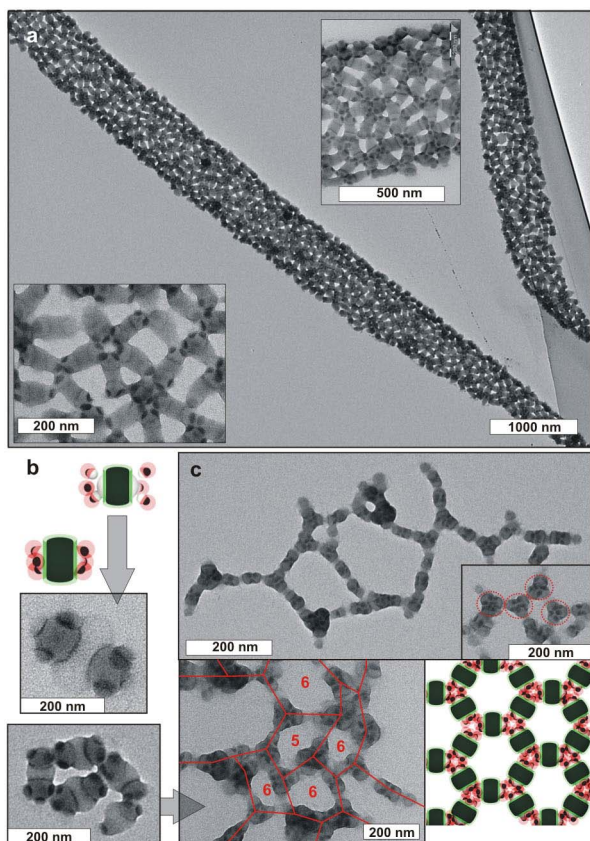
polymers for varying SD<sup>MS</sup> to l-SB<sup>M</sup> mixing ratios of 40:2, 20:2, 10:2 and 5:2. Scale bars, 5  $\mu\text{m}$ . We count segments and multiply the resulting average value by the average segment length,  $L_n = 55$  nm, to yield the average length of the co-assembly. Data are plotted against the normalized frequency and fitted by a Schulz-Flory frequency distribution (squares),  $F(L) = \exp(-L/L_n)/L_n$ , showing a continuous decrease in the average length coinciding with the mixing ratio. The similar timescales of onset of aggregation combined with matching sizes of CBBs both promote efficient end-capping.



**Extended Data Figure 8 | Stability of  $[\text{SD}^{\text{M}}\text{S}]_m$  colloidal polymer versus  $\text{SB}^{\text{T}}/[\text{SD}^{\text{M}}\text{S}]_m$  co-assembly in polar solvents.** Photograph of two colloidal solutions after dialysis into ethanol. Left: instant precipitation of the

$[\text{SD}^{\text{M}}\text{S}]_m$  colloidal homopolymer as a result of complete collapse of the M corona. Right: the  $\text{SB}^{\text{T}}/[\text{SD}^{\text{M}}\text{S}]_m$  colloidal co-assembly is stabilized by the soluble T corona patches. Precipitation is not observed even after weeks.





#### Extended Data Figure 9 | Colloidal molecules and two-dimensional network formation.

**a**, Co-assembly of four  $\text{SB}^{\text{M}}$  with one  $\text{SD}^{\text{T}}\text{S}$  CBB by means of collapsing S patches in acetone/propan-2-ol (60:40 v/v) yields colloidal molecules with one or two end-caps of the  $(\text{SB}^{\text{M}})_2-(\text{SD}^{\text{T}}\text{S})-(\text{SB}^{\text{M}})_2$  type, permitting three or four nearest neighbours at every linking point in large-area networks. **b**, Co-assembly of six  $\text{SB}^{\text{M}}$  with one  $\text{SD}^{\text{T}}\text{S}$  CBB by means of collapsing S patches in acetone/propan-2-ol (60:40 v/v) yields colloidal molecules of the  $(\text{SB}^{\text{M}})_3-(\text{SD}^{\text{T}}\text{S})-(\text{SB}^{\text{M}})_3$  type. **c**, Addition of propan-2-ol until acetone/propan-2-ol reached 20:80 (v/v) selectively collapses the terminal M corona, allowing two nearest neighbours at every linking point of the network in solution. On some occasions the network takes on the form of a distorted kagome lattice.

Extended Data Table 1 | Characteristics of triblock terpolymers and CBBs

CBB Code*	Triblock Terpolymer†	$\frac{V_A}{V_B}‡$	$r_C = \frac{V_C}{V_A + V_B}‡$	$d_{CBB}§$ [nm]	$R_{B-core}  $ [nm]	$V_B¶$ [x10 <sup>3</sup> nm <sup>3</sup> ]	$N_{agg}^\#$	$V_S^{☆}$ [x10 <sup>3</sup> nm <sup>3</sup> ]
<b>AB<sup>C</sup>A CBBs</b>								
<b>SD<sup>M</sup>S</b>	S <sub>560</sub> D <sub>580</sub> M <sub>540</sub> <sup>195</sup>	0.37	0.33	55x24	55x30	71.2	370	26.3
<b>SD<sup>T</sup>S</b>	S <sub>510</sub> D <sub>540</sub> T <sub>460</sub> <sup>258</sup>	0.36	0.44	60x30	57x38	96.1	385	34.6
<b>SB<sup>M</sup>S</b>	S <sub>590</sub> B <sub>2700</sub> M <sub>560</sub> <sup>265</sup>	0.38	0.25	100x31	100x50	394.0	2300	149.7
<b>s-AB<sup>C(y)</sup> CBBs of similar size, yet decreasing corona length (l), (m), (s)</b>								
<b>s-SB<sup>M(l)</sup></b>	S <sub>310</sub> B <sub>150</sub> M <sub>340</sub> <sup>74</sup>	3.57	0.74	19	5 ± 2	0.5	40	1.9
<b>s-SB<sup>M(m)</sup></b>	S <sub>510</sub> B <sub>260</sub> M <sub>260</sub> <sup>90</sup>	4.20	0.34	28	7 ± 2	1.4	60	5.9
<b>s-SB<sup>M(s)</sup></b>	S <sub>700</sub> B <sub>270</sub> M <sub>180</sub> <sup>105</sup>	4.50	0.15	29	7 ± 2	1.4	57	6.3
<b>x-AB<sup>C</sup> CBBs of different overall size s-, m-, l-.</b>								
<b>s-SB<sup>M</sup></b>	S <sub>310</sub> B <sub>150</sub> M <sub>340</sub> <sup>74</sup>	3.57	0.74	19	5 ± 2	0.5	40	1.9
<b>m-SB<sup>M</sup></b>	S <sub>560</sub> B <sub>580</sub> M <sub>540</sub> <sup>143</sup>	1.78	0.88	35	11 ± 2	5.6	140	9.9
<b>l-SB<sup>M</sup></b>	S <sub>610</sub> B <sub>640</sub> M <sub>290</sub> <sup>140</sup>	1.72	0.26	47	15 ± 2	14.1	260	24.3
<b>AB CBBs with different nature of the corona</b>								
<b>SB<sup>T</sup></b>	S <sub>510</sub> B <sub>540</sub> T <sub>480</sub> <sup>155</sup>	1.70	0.45	28	9 ± 2	3.1	60	5.2
<b>SB<sup>V</sup></b>	S <sub>360</sub> B <sub>380</sub> V <sub>590</sub> <sup>120</sup>	1.67	0.95	22	7 ± 2	1.4	54	2.4

\* The syntheses and characterization of SBM, SBV and SBT triblock terpolymers were reported in detail previously<sup>31–35</sup>. Superscripts denote the size of the soluble corona patches in relation to the other CBBs: s, small; m, medium; l, large.

† Subscripts denote the degrees of polymerization of the corresponding blocks, and the superscript is the molecular mass in kg mol<sup>−1</sup> determined with combined <sup>1</sup>H-NMR and GPC (polydispersity index < 1.15) measurements.

‡ Volume fractions  $V_A$ ,  $V_B$  and  $V_C$  were calculated from molar volumes and degrees of polymerization.

§ For calculations see Methods. The diameter ( $d$ ) and width ( $w$ ) of cylindrical segments of AB<sup>C</sup>A units and (AB<sup>C</sup>A)<sub>x</sub> segments are average values from TEM image analysis.

|| Average of 250 measured particle core radii in TEM image analysis.

¶ Particle core volumes calculated from measured core radii.

# Average aggregation number of polymer chains per particle.

☆ Calculated by applying the relation  $V_S/V_B$ .

# Evidence for high salinity of Early Cretaceous sea water from the Chesapeake Bay crater

Ward E. Sanford<sup>1</sup>, Michael W. Dougherty<sup>1</sup>, Tyler B. Coplen<sup>1</sup>, Andrew G. Hunt<sup>2</sup> & Thomas D. Bullen<sup>3</sup>

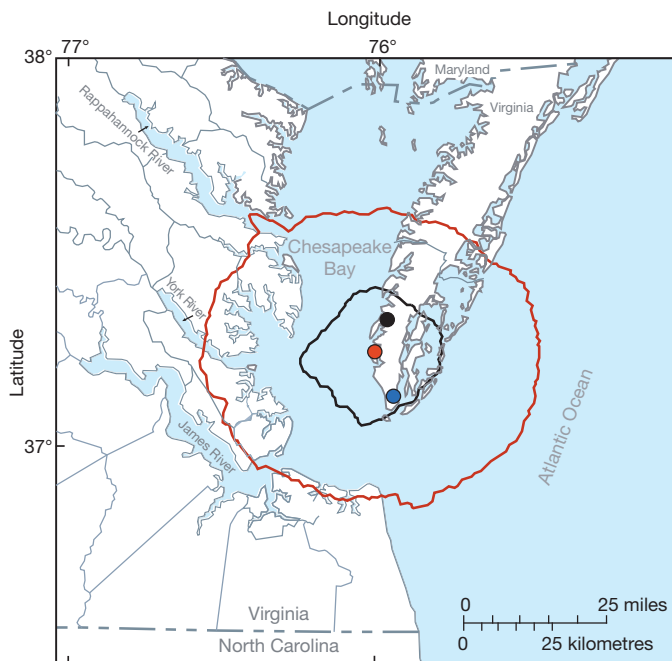
High-salinity groundwater more than 1,000 metres deep in the Atlantic coastal plain of the USA has been documented in several locations<sup>1,2</sup>, most recently within the 35-million-year-old Chesapeake Bay impact crater<sup>3–5</sup>. Suggestions for the origin of increased salinity in the crater have included evaporite dissolution<sup>6</sup>, osmosis<sup>6</sup> and evaporation from heating<sup>7</sup> associated with the bolide impact. Here we present chemical, isotopic and physical evidence that together indicate that groundwater in the Chesapeake crater is remnant Early Cretaceous North Atlantic (ECNA) sea water. We find that the sea water is probably 100–145 million years old and that it has an average salinity of about 70 per mil, which is twice that of modern sea water and consistent with the nearly closed ECNA basin<sup>8</sup>. Previous evidence for temperature and salinity levels of ancient oceans have been estimated indirectly from geochemical, isotopic and palaeontological analyses of solid materials in deep sediment cores. In contrast, our study identifies ancient sea water *in situ* and provides a direct estimate of its age and salinity. Moreover, we suggest that it is likely that remnants of ECNA sea water persist in deep sediments at many locations along the Atlantic margin.

Hypersaline groundwater in the deep Atlantic Coastal Plain of North America was discovered in several locations decades ago. Off the coast of Georgia, hypersaline brines ( $>100 \text{ g kg}^{-1}$ ) associated with anhydrite deposits exist at depths exceeding 1.2 km in Lower Cretaceous strata<sup>1</sup>. On Cape Hatteras, North Carolina, drilling in 1946 recovered brine from a depth of 2.0 km with a salinity of  $117 \text{ g kg}^{-1}$  (ref. 9). In southeastern Maryland, actual and estimated chloride concentrations of  $40 \text{ g kg}^{-1}$  and  $50 \text{ g kg}^{-1}$  were reported<sup>2,10</sup> at 1.0 km and 1.2 km. In southern New Jersey, at 1.0 km, a chloride concentration was observed at  $25 \text{ g kg}^{-1}$  (ref. 2). Observations have also indicated that the position of the freshwater–saltwater transition zone extends tens of kilometres farther inland near the southernmost Chesapeake Bay<sup>11</sup>. This ‘inland wedge’ was eventually recognized as being collocated with the Chesapeake Bay impact structure<sup>4,6</sup>, and hypersaline water was detected there at Kiptopeake, Virginia, at a depth of 0.6 km (Fig. 1)<sup>4</sup>. Two original theories for the origin of the hypersalinity at the crater—evaporite dissolution and reverse osmosis<sup>6</sup>—have now been dismissed because of the lack of proximal evaporites and the necessary hydrogeologic conditions, respectively<sup>12</sup>.

The Chesapeake Bay impact structure was created by a bolide impact 35 Myr ago in the Late Eocene epoch<sup>3,6</sup>. Deep drilling by the US Geological Survey (USGS) at Cape Charles, Virginia<sup>13,14</sup>, and the USGS and the International Continental Drilling Program (ICDP) at Eyreville Neck, Virginia<sup>15,16</sup>, recovered core and revealed a stratigraphic section documenting the sediment and bedrock disruption associated with the ‘wet-target’ (that is, the location had standing water, usually the ocean) impact. The test hole at Cape Charles<sup>14</sup>, drilled in 2004, revealed post-impact sediments underlain by synimpact (that is, formed by the impact) sediment-clast resurge, crystalline-clast breccias and suevites down to the bottom of the hole at 823 m. Wells at this site were installed at depths of 418 m and 692 m. The deep hole at Eyreville Neck, drilled in 2005, revealed post-impact sediments down

to 444 m, sediment-clast resurge breccias down to 1,096 m, a granite megablock down to 1,371 m, suevites, cataclasites and crystalline-clast breccias down to 1,551 m, and schist and pegmatite allochthonous blocks down to 1,766 m (Fig. 2a). Drilling did not reach the bottom of the excavated crater. The drill-stem casing was left in the hole down to 1,330 m, which allowed for later sampling and chemical analysis of formation water.

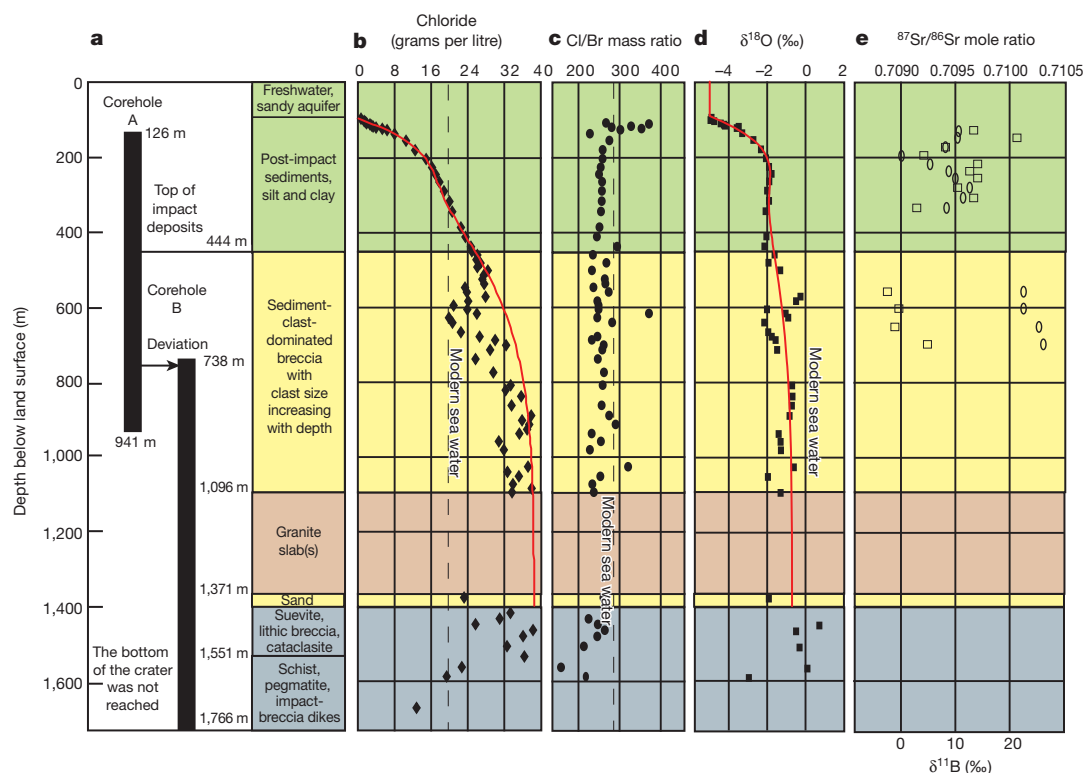
Before data were retrieved from deep coring, impact heat and associated evaporation was suggested as a source of the elevated salinity. Simulations of the hydrothermal response to the impact suggested that hydrothermal and vapour phase conditions could exist during and following the impact<sup>17</sup>. To investigate the thermal history of the crater and overlying sediments, vitrinite reflectance was measured on sediment cores from the Eyreville Neck and Cape Charles sites<sup>18</sup>. Results showed a thermal signature of upward movement of groundwater in the 10,000 years or so following the impact, but a heat balance based on the suevite temperature and the quantity of hypersaline groundwater in the crater demonstrated that there was inadequate heat to account for the excess salinity solely by steam production and loss<sup>19</sup>.



**Figure 1 | Map showing locations of Chesapeake Bay crater and coreholes.** Map showing location of the Chesapeake Bay impact structure and sites of various deep coreholes and wells<sup>19</sup>. The black and red circular lines mark the outer edges of the inner and outer craters. The black circle marks the location of the ICDP-USGS deep corehole and well<sup>15,16</sup>. The red circle marks the location of the USGS corehole and wells at Cape Charles<sup>13,14</sup>. The blue circle marks the location of the USGS corehole and wells at Kiptopeake<sup>4,5</sup>.

<sup>1</sup>US Geological Survey, Mail Stop 431, Reston, Virginia 20192, USA. <sup>2</sup>US Geological Survey, Federal Center, Box 25046, Mail Stop 964, Denver, Colorado 80225, USA. <sup>3</sup>US Geological Survey, McKelvey Building 15, 345 Middlefield Road, Mail Stop 420, Menlo Park, California 94025, USA.

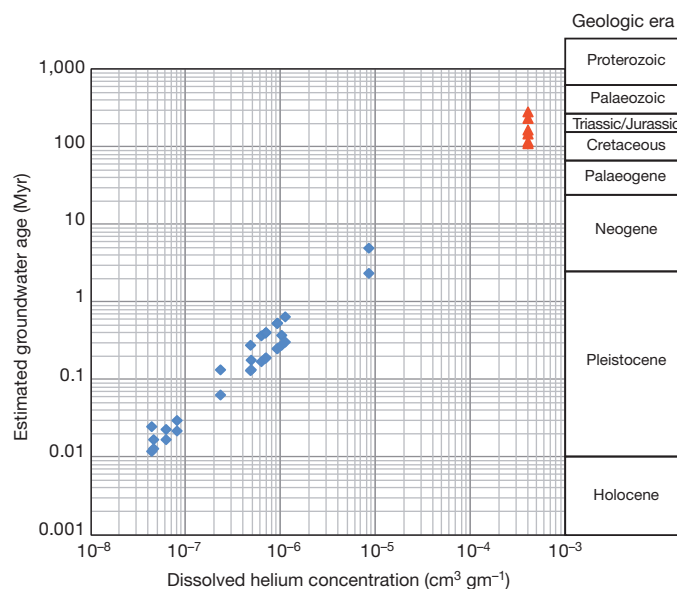




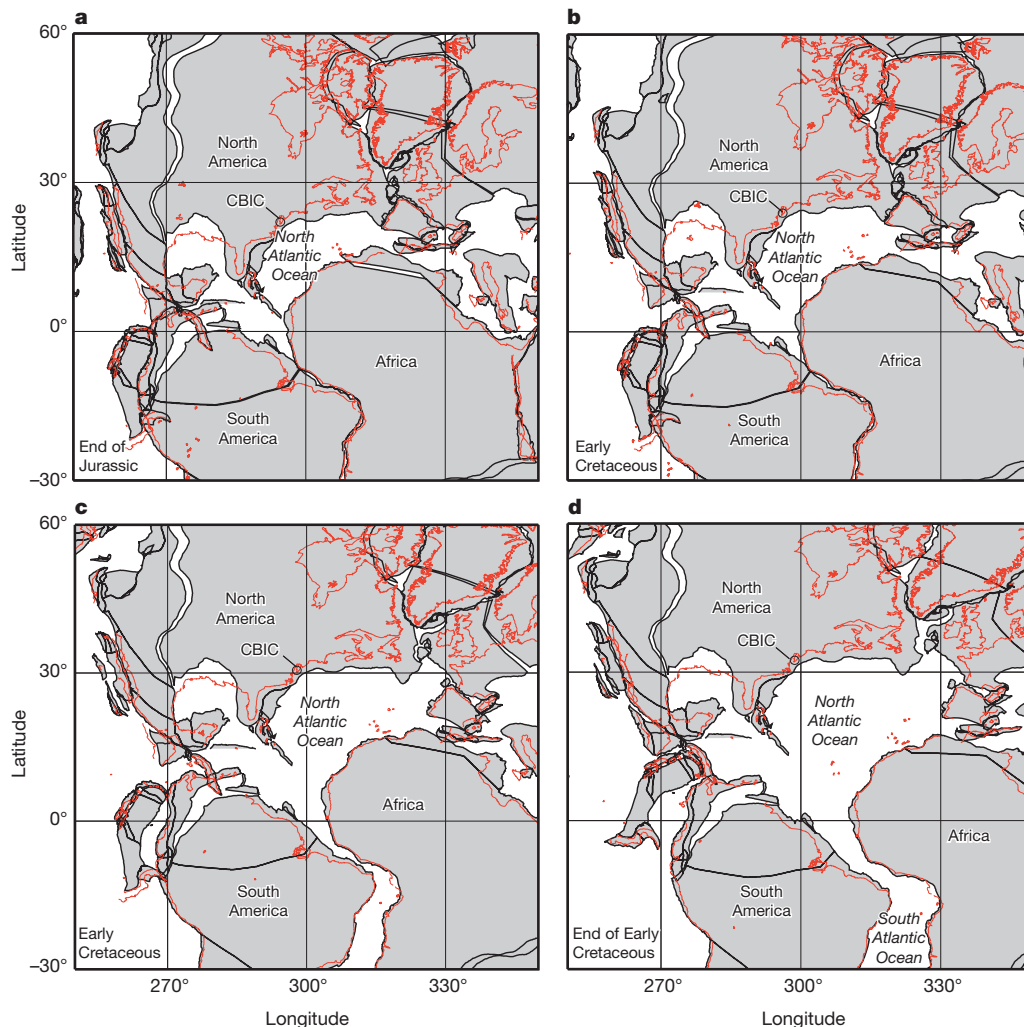
**Figure 2 | Chloride and isotope results from the deep ICDP-USGS corehole.** Results from the ICDP-USGS corehole in the Chesapeake Bay impact crater near Eyreville, Virginia, including drilling depths and sediment types recovered (a), dissolved chloride concentrations (b), chloride/bromide mass ratios (c), oxygen isotope measurements of the pore water (d) and dissolved strontium and dissolved boron isotope measurements (e). Open circles are  $^{87}\text{Sr}/^{86}\text{Sr}$  mole ratio results, open squares are boron isotope results. The solid

red curved lines are simulation results assuming values of molecular diffusion  $D_m$  and upward fluid flux  $q$  from compaction of  $2 \times 10^{-11} \text{ m}^2$  and  $3.5 \text{ m Myr}^{-1}$ , respectively. Additional details of the simulation methods can be found in Methods and Supplementary Information. Each water sample was analysed only once, and so in all figures each point on the graph represents a single analysis.

Pore water was extracted from cores at the Eyreville site over the entire 1,766 m. Ten-cm-long core samples were taken approximately every 10–20 m, and water was extracted from the core by high-speed centrifuge. Analyses were performed to determine the concentrations of major cations and anions (Extended Data Fig. 1), pH, alkalinity, and the isotopic compositions of hydrogen ( $\delta^2\text{H}$ ) and oxygen ( $\delta^{18}\text{O}$ ) of water, dissolved boron ( $\delta^{11}\text{B}$ ) and strontium ( $^{87}\text{Sr}/^{86}\text{Sr}$ ), and sulphur ( $\delta^{34}\text{S}$ ) and oxygen ( $\delta^{18}\text{O}$ ) of dissolved sulphate<sup>19</sup>. The dissolved chloride concentration at the site increases with depth continuously through the post-impact section, surpassing the seawater concentration well before the top of the crater-fill deposits (Fig. 2b). Concentrations continue to increase erratically with depth in the sediment-clast breccia until reaching approximately 38 g per litre (twice that of modern sea water) by 900 m. The Cl/Br mass ratio in pore waters at the site is relatively constant through the entire section, at a mean value of 262 (that in modern sea water is 292) (Fig. 2c). This rules out halite dissolution as a cause of elevated salinity<sup>20</sup> and is indicative rather of an ancient sea water that is slightly depleted in chloride relative to bromide, consistent with the vast halite deposits created during the Jurassic and Cretaceous periods in the Gulf of Mexico and South Atlantic Basins. The  $\delta^{18}\text{O}$  values of the pore water at the site increase with depth (Fig. 2d) in a pattern similar to that of chloride. Strontium isotope ratios are known to have varied substantially in sea water since the Mesozoic era (Early Cretaceous  $^{87}\text{Sr}/^{86}\text{Sr}$  mole ratios varied between 0.7073 and 0.7075)<sup>21</sup>, and thus might be a good indicator of ancient sea water, but samples analysed here reflect leaching of strontium<sup>22</sup> from continental-sourced Potomac Formation sediment clasts hosting the water (Fig. 2e). Boron isotopic compositions also suggest a strong water–rock interaction, because  $\delta^{11}\text{B}$  values differ substantially from sea water (11‰) (Fig. 2e) and are consistent with movement through and interaction with



**Figure 3 | Helium-based age estimates for the deep Chesapeake crater groundwater.** Estimated groundwater ages in the mid-Atlantic coastal plain of North America, based on dissolved helium concentrations. Blue diamonds represent estimates of groundwater ages from wells sampled in the coastal plain sediments of Maryland<sup>25</sup>. These ages are based on  $^4\text{He}$ -accumulation rates calibrated against  $^{14}\text{C}$ -based and  $^{36}\text{Cl}$ -based ages calculated for a set of wells in the same geologic formation as the crater wells<sup>25</sup>. Red triangles represent estimates of age for the groundwater in the Chesapeake Bay impact crater sampled at the well near Eyreville, Virginia. These latter age estimates are a linear extrapolation using the  $^4\text{He}$ -accumulation rates calculated from the Maryland samples.



**Figure 4 | Plate-tectonic reconstruction of the ECNA ocean.** Plate-tectonic reconstruction of the North Atlantic Ocean through the Early Cretaceous epoch<sup>8</sup> at 145 Myr ago (a), 130 Myr ago (b), 115 Myr ago (c) and 100 Myr ago (d). Red lines represent modern coastlines. Grey shapes represent continental plates and fragments. The open circle on the Atlantic coast of North America is the site of the late Eocene (35-Myr-old) Chesapeake Bay impact

crater (CBIC). The maps are an attempt to delineate the edge of the continents, and do not show the continental regions that were covered by ocean water during the global Cretaceous transgression. They were constructed from a software program located at <http://www.odsn.de/odsn/services/paleomap/paleomap.html>.

igneous rocks in the crater and/or pre-impact crystalline basement (Extended Data Fig. 2).

The curvature of the chloride and  $\delta^{18}\text{O}$  profiles suggests diffusion and slow upward advection as controlling processes<sup>23</sup>. One-dimensional transport simulations were carried out to test this hypothesis. The advection–diffusion equation was solved with an upper boundary condition that allowed for post-impact sequences of sediment to be deposited over time and for the upper boundary salinity to alternate between that of fresh water and salt water, depending on the inferred position of the coastline from the known depositional environments and hiatuses (Extended Data Fig. 3). The simulated chloride values could be made to match the observed values very closely in the upper 500 m of the section (Fig. 2) when a diffusion coefficient  $D_m$  of  $2 \times 10^{-11} \text{ m}^2 \text{ s}^{-1}$  was used along with an upward fluid flux  $q$  of  $3.5 \text{ m Myr}^{-1}$  and original chloride concentrations of 38 g per litre in the sediment-clast breccia section. Observed variable concentrations in the sediment-clast breccias lower than those simulated are attributed to clay clasts that are retaining brackish water<sup>19</sup>. The  $\delta^{18}\text{O}$  profile was also simulated with the same diffusion coefficient and upward fluid flux, and initial  $\delta^{18}\text{O}$  values (relative to the Vienna Standard Mean Ocean Water (VSMOW)–Standard Light Antarctic Precipitation (SLAP) standards) of  $-1.0\text{‰}$  and  $-0.8\text{‰}$  in the coastal plain and sediment-clast breccias. The simulated

$\delta^{18}\text{O}$  values also make a close match with the observed data down to 500 m. Observed and simulated  $\delta^{37}\text{Cl}$  values suggest possible isotopic effects related to the impact, but are also consistent with an ECNA origin for the pore water (Extended Data Fig. 4).

The minor inflections in the Cl and  $\delta^{18}\text{O}$  data can be replicated only by including sedimentation and temporal salinity changes at the upper boundary. The overall goodness-of-fit to these simulations (Extended Data Fig. 5) is evidence that the crater fill has been undisturbed by lateral groundwater movement since the time of impact, with the solutes being influenced only by diffusion and advection consistent with burial and compaction of fine-grained sediment<sup>23</sup>. The initial condition of 38 g per litre for the chloride concentration simulation represents water that was trapped in and below the coastal plain sediments before the bolide impact occurred. Simulation of chloride concentrations in the pre-impact sediments confirms that most of the pore water would have remained undisturbed between the end of the Early Cretaceous period and the time of impact (Extended Data Fig. 6). The impact disrupted the sediment and basement crystalline rocks, along with their pore water, redepositing them in the crater. Because much of the crater fill is composed of clasts many metres in diameter, most of the initial pore water would have been retained through the collapse and resurge phases after impact<sup>16</sup>. Vertical pore-fluid flux due to loose-sediment

compaction immediately following the resurge was estimated to be around  $30 \text{ m kyr}^{-1}$  (ref. 18). But the much slower  $3.5 \text{ m Myr}^{-1}$  flux over the intervening 35 Myr indicated by the chloride and  $\delta^{18}\text{O}$  profile is consistent with loading rates during post-impact sedimentation.

Dissolved gas samples were collected from the deep wells at Cape Charles and Eyreville. At Eyreville, the well casing in corehole B was left *in situ* at the very top of the granite megablock(s). Helium concentrations and isotope ratios ( $R/R_a$ ) for deep Cape Charles and Eyreville groundwater were found to be  $5.0 \times 10^{-3} \text{ cm}^3 \text{ g}^{-1}$  at STP (standard temperature and pressure) and  $4.0 \times 10^{-3} \text{ cm}^3 \text{ g}^{-1}$  at STP, and 0.44 and 0.04, respectively. The low  $R/R_a$  number at Eyreville suggests that this sample's helium concentration represents predominantly *in situ* helium accumulation over time<sup>24</sup>, and so we used that concentration to calculate a He-based groundwater age. From recent estimates of  $^4\text{He}$  accumulation rates in the coastal plain sediments<sup>25</sup>, the helium-age of the saline groundwater would be  $180 \pm 80 \text{ Myr}$  (Fig. 3) given the variability in accumulation rates and the extrapolation to the much older age. The age can be further constrained to the Early Cretaceous (145–100 Myr ago), because the Lower Cretaceous Potomac Formation is the oldest local sediment, resting directly on the crystalline basement under the coastal plain (pages 19 and 20 in ref. 26). Given that the depositional environment of the Lower Cretaceous Potomac Formation is coastal fluvial and deltaic, the water in the Lower Cretaceous Potomac Formation and the upper crystalline basement would have been replaced with Early Cretaceous North Atlantic (ECNA) sea water during the global Cretaceous transgression.

For the North Atlantic Ocean, the Cretaceous period was a time of transition from a closed rift basin to an open ocean (Fig. 4). During the Jurassic period, the early North Atlantic must have had a very high salinity, evidenced by the fact that it was depositing salt as evaporite deposits around its border<sup>27</sup>. Rifting that would lead to the South Atlantic Ocean was only beginning during the Early Cretaceous. As small openings to the ECNA waxed and waned and eventually became wider during the Early Cretaceous, the salinity would have fluctuated but slowly declined overall from highly saline, eventually reaching a near-modern value. Anoxic events recorded in the ECNA suggest a lack of basin-wide circulation<sup>28</sup>, and some  $\delta^{18}\text{O}$  evidence from foraminifera in cores suggest a salinity higher than that of modern sea water<sup>29,30</sup>. By the end of the Early Cretaceous, the openings, especially to the South Atlantic, were becoming substantially wider (Fig. 4d). Even a small opening could lead to substantial dilution of salinity. As a modern example, the only opening to the Mediterranean Sea is the Strait of Gibraltar, yet its salinity (38‰) is only about 10% above that of modern ocean salinity. By the beginning of the Late Cretaceous (100 Myr ago), North Atlantic salinity was probably near that of modern sea water.

We conclude that the saline water currently present in the Chesapeake Bay impact crater's resurge breccia and crystalline blocks is a remnant of ECNA sea water, and that the ECNA had an average salinity of about 70‰, or twice that of modern sea water. In addition, given that several other locations along the North American Atlantic Coastal Plain have salinities greater than that of sea water at depth, we believe remnants of ECNA sea water probably persist at many other locations along the Atlantic margin as well.

## METHODS SUMMARY

At the Eyreville ICPD-USGS site, 10-cm-long sections of core were removed every 10–20 m for pore-water analysis. Most of the unconsolidated cores removed for pore-water analysis were bisected, and one half was returned for archival purposes. A few millimetres of the outer rind of each sample were removed with a knife to remove visible drilling mud from the outside of the core and any that may have invaded the outer edge of the core. Core sections were then placed in glass jars that had been flushed with nitrogen, sealed and refrigerated at  $10^\circ\text{C}$ .

The core samples were placed in one of two USGS high-speed centrifuges in Reston, Virginia, and Denver, Colorado, to extract pore water. Once extracted, the pore-water samples were analysed for pH, alkalinity and specific conductance<sup>19</sup> by using Hach small-volume test kits. The remaining water was then separated into

vials for various analyses and refrigerated. Analyses for cations and anions and the stable isotopes of water were analysed in USGS laboratories in Reston, Virginia, by ion chromatography and inductively coupled plasma-optical spectroscopy, respectively. Hydrogen and oxygen isotopes of water were measured, respectively, by gaseous hydrogen and carbon dioxide equilibration and dual-inlet isotope-ratio mass spectrometry. Strontium and boron isotope ratios were measured in a USGS laboratory in Menlo Park, California, by inductively coupled plasma mass spectrometry.

For helium in the deep wells, water samples were retrieved by using a 6-litre, stainless steel Kemmerer sampling flask attached to a portable cable winch fitted with a digital depth counter. After being lowered to near the bottom of the well casing, the Kemmerer flask was closed by a sender weight and retrieved. Dissolved helium and helium isotope ratios were measured in the USGS Noble Gas Laboratory in Denver, Colorado, using a capacitance manometer, and a MAP-215-50 mass spectrometer, respectively.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 March; accepted 19 September 2013.

- Manheim, F. & Paull, C. Patterns of groundwater salinity changes in a deep continental-oceanic transect off the southeastern Atlantic coast of the USA. *J. Hydrol.* **54**, 95–105 (1981).
- Meisler, H. The occurrence and geochemistry of salty groundwater in the Northern Atlantic coastal plain. *US Geol. Surv. Prof. Pap.* Report 1404-D (1989).
- Poag, C., Koeberl, C. & Reimold, W. *The Chesapeake Crater—Geology and Geophysics of a Late Eocene Submarine Impact Structure* (Springer, 2003).
- McFarland, E. & Bruce, T. Distribution, origin, and resource-management implications of ground-water salinity along the western margin of the Chesapeake Bay impact structure in eastern Virginia. *US Geol. Surv. Prof. Pap.* **1688**, K1–K32 (2005).
- McFarland, E. Groundwater quality data and regional trends in the Virginia Coastal Plain, 1906–2007. *US Geol. Surv. Prof. Pap.* Report **1772** (2010).
- Poag, C. *Chesapeake Invader: Discovering America's Giant Meteorite Crater* (Princeton Univ. Press, 1999).
- Sanford, W. Heat flow and brine evolution following the Chesapeake Bay bolide impact. *J. Geochem. Explor.* **78–79**, 243–247 (2003).
- Hay, W. *et al.* in *The Evolution of Cretaceous Ocean/Climate Systems* (eds Barrera, E. & Johnson, E.) *Geol. Soc. Am. Spec. Pap.* **332**, 1–47 (1999).
- Spangler, W. Subsurface geology of Atlantic Coastal Plain of North Carolina. *Am. Assoc. Petrol. Geol. Bull.* **34**, 100–132 (1950).
- Radford, L., Cobb, L. & McCoy, R. *Atlantic Coastal Plain Geothermal Drilling Program, DOE/Crisfield Airport No. 1 Well, Somerset County, Maryland*. Tech. Rep. ET/28373–1 (Department of Energy, 1980).
- Cederstrom, D. Chloride in groundwater in the coastal plain of Virginia. *Virginia Geol. Surv. Bull.* **58**, 1–384 (1943).
- Neuzil, C. & Provost, A. Recent experimental data may point to a greater role for osmotic pressures in the subsurface. *Wat. Resour. Res.* **45**, 1–14 (2009).
- Sanford, W. *et al.* Drilling the central crater of the Chesapeake Bay impact structure: a first look. *Eos* **85**, 369–377 (2004).
- Gohn, G. *et al.* Site report for USGS test holes drilled at Cape Charles, Northampton county, Virginia, in 2004. *US Geol. Surv. Open-file Rep.* Report 2007–1094 (2007).
- Gohn, G. *et al.* Chesapeake Bay impact structure drilled. *Eos* **87**, 349–355 (2006).
- Gohn, G. *et al.* Deep drilling into the Chesapeake Bay impact structure. *Science* **320**, 1740–1745 (2008).
- Sanford, W. A simulation of the hydrothermal response to the Chesapeake Bay bolide impact. *Geofluids* **5**, 185–201 (2005).
- Malinconico, M., Sanford, W. & Horton, J. Postimpact heat conduction and compaction-driven fluid flow in the Chesapeake Bay impact structure based on downhole vitrinite reflectance data, ICDP-USGS Eyreville deep core holes and Cape Charles test holes. *Spec. Pap. Geol. Soc. Am.* **458**, 905–930 (2009).
- Sanford, W. *et al.* Pore-water chemistry from the ICDP-USGS core hole in the Chesapeake Bay impact structure—implications for paleohydrology, microbial habitat, and water resources. *Spec. Pap. Geol. Soc. Am.* **458**, 867–890 (2009).
- Davis, S., Whittemore, D. & Fabryka-Martin, J. Uses of chloride/bromide ratios in studies of potable water. *Ground Water* **36**, 338–350 (1998).
- Jones, C., Jenkins, H., Coe, A. & Stephen, H. Strontium isotope variations in Jurassic and Cretaceous seawater. *Geochim. Cosmochim. Acta* **58**, 3061–3074 (1994).
- Bullen, T., Krabbenhoft, D. & Kendall, C. Kinetic and mineralogical controls on the evolution of groundwater chemistry and  $^{87}\text{Sr}/^{86}\text{Sr}$  in a sandy silicate aquifer, northern Wisconsin, USA. *Geochim. Cosmochim. Acta* **60**, 1807–1821 (1996).
- Ingebritsen, S., Sanford, W. & Neuzil, C. *Groundwater in Geologic Process* (Cambridge Univ. Press, 2006).
- Oxburgh, E., O'Nions, R. & Hill, R. Helium isotopes in sedimentary basins. *Nature* **324**, 632–635 (1986).
- Plummer, L. *et al.* Old groundwater in parts of the upper Patapsco aquifer, Atlantic Coastal Plain, Maryland, USA: evidence from radiocarbon, chlorine-36 and helium-4. *Hydrogeol. J.* **20**, 1269–1294 (2012).
- McFarland, E. & Bruce, T. The Virginia coastal plain hydrogeologic framework. *US Geol. Surv. Prof. Pap.* Report 1731 (2006).
- Evans, R. Origin and significance of evaporites in basins around Atlantic margin. *Am. Assoc. Petrol. Geol. Bull.* **62**, 223–234 (1978).



28. Arthur, M. & Natland, J. in *Deep Drilling in the Atlantic Ocean: Continental Margins and Paleoenvironment* Vol. 3 (eds Talawani, M., Hay, W. & Ryan, B. F.) 375–401 (American Geophysical Union, 1979).
29. Friedrich, O., Erbacher, J., Moriya, K. & Wilson, P. Warm saline intermediate waters in the Cretaceous tropical Atlantic Ocean. *Nature Geosci.* **1**, 453–457 (2008).
30. Wagner, T. *et al.* Rapid warming and salinity changes of Cretaceous surface waters in the subtropical North Atlantic. *Geology* **36**, 203–206 (2008).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** H. Qi of the USGS in Reston assisted with  $\delta^2\text{H}$ ,  $\delta^{18}\text{O}$ , and  $\delta^{34}\text{S}$  measurements of water and dissolved sulphate. N. Sturchio of the University of Chicago analysed  $\delta^{37}\text{Cl}$ . D. Webster and N. Bach of the USGS extracted pore waters from the cores in Reston, Virginia. J. Little, C. Johnson-Griscavage and J. Koch of the USGS extracted pore waters from the cores in Denver, Colorado. G. Casile of the USGS in Reston sampled the deep wells for dissolved gas analysis at Eyreville and Cape Charles. W. Aleman-Gonzalez, N. Bach, C. Durand, L. Edwards, J. Glidewell, J. Kirshtein,

T. Kraemer, D. Larsen, M. Lowit, H. Michael, J. Murray, A. Palmer-Julson, E. Seefelt, J. Self-Trail, M. Voytek, D. Webster and B. Zinn all helped collect core samples during the three months of 24-hour drilling. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the US government.

**Author Contributions** W.E.S. compiled and interpreted the data, performed the transport simulations and wrote this article. M.W.D. analysed pore waters for the major anions and cations. T.B.C. analysed the pore waters and dissolved sulphate for their  $\delta^2\text{H}$ ,  $\delta^{18}\text{O}$  and  $\delta^{34}\text{S}$  values. A.G.H. analysed the well-water samples for the total dissolved helium and helium isotope ratios. T.D.B. analysed the pore waters for the boron and strontium isotopic compositions and interpreted their values.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.E.S. ([wsanford@usgs.gov](mailto:wsanford@usgs.gov)).

## METHODS

**Core sample collection.** At the Eyreville ICPD-USGS site, 10-cm-long sections of core were removed every 10–20 m for pore-water analysis. Cores from core hole A were 8.5 cm in diameter from 126 m to 591 m and 6.35 cm from 592 m to 941 m, whereas those from core hole B were 6.35 cm from 738 m to 1,101 m and 4.76 cm from 1,101 m to 1,706 m. Most of the unconsolidated cores removed for pore-water analysis were bisected, and one half was returned for archival purposes. All core split samples were collected before the cores were washed. A few millimetres of the outer rind of each sample were removed with a knife to remove visible drilling mud from the outside of the core and any that might have invaded the outer edge of the core. Core sections were then placed in glass jars that had been flushed with nitrogen, sealed and refrigerated at 10 °C. The jarred samples were then transported and stored at a USGS laboratory in Reston at 5 °C until pore-water extraction.

**Core sample extraction.** The core samples were placed in one of two USGS high-speed centrifuges in Reston, Virginia, and Denver, Colorado, to extract pore water. Samples were removed from the jars and pulverized, and subsamples were spun at room temperature at 40,000 revolutions per minute (r.p.m.) for about 30 min. Once collected, the pore-water samples were analysed for pH, alkalinity and specific conductance<sup>19</sup> by using Hach small-volume test kits. The remaining water was then separated into vials for various analyses and refrigerated. Water and solute analyses included major anions and cations,  $\delta^2\text{H}$  and  $\delta^{18}\text{O}$  of water,  $^{87}\text{Sr}/^{86}\text{Sr}$  of dissolved strontium, dissolved organic carbon, and  $\delta^{34}\text{S}$  and  $\delta^{18}\text{O}$  of sulphate<sup>19</sup>. To ensure the samples were not contaminated by drilling fluids, selected samples were also analysed for polar-organic compounds that were characteristic of the drilling fluids<sup>31</sup>. Drilling fluids were not detected in any of the samples.

**Dissolved ion and isotope methods.** Analyses for cations and anions and the stable isotopes of water were performed in USGS laboratories in Reston, Virginia. The anions and cations were analysed by ion chromatography and inductively coupled plasma-optical spectroscopy, respectively. Hydrogen and oxygen isotopes of water were measured, respectively, by gaseous hydrogen<sup>32</sup> and carbon dioxide equilibration<sup>33</sup> and dual-inlet isotope-ratio mass spectrometry. Sulphur and oxygen isotopes of dissolved sulphate were analysed by conversion of sulphate to barium sulphate<sup>34</sup> and analysis by continuous-flow isotope-ratio mass spectrometry<sup>35</sup>. Strontium and boron isotopes were measured in a USGS laboratory in Menlo Park, California, by inductively thermal ionization mass spectrometry<sup>22</sup>, and negative thermal ionization mass spectrometry<sup>36</sup>, respectively. Chloride stable isotopes were analysed at the University of Chicago by continuous-flow isotope-ratio mass spectrometry<sup>37</sup>.

**Helium sample collection.** In the deep wells at both Eyreville and Cape Charles, water samples were retrieved by using a 6-litre, stainless steel Kemmerer sampling flask attached to a portable cable winch fitted with a digital depth counter. The deep wells were first purged to ensure that pristine formation water was in the well casing. After being lowered to near the bottom of the well casing, the Kemmerer flask was closed by a sender weight and retrieved. The dissolved gas samples were then transferred to a 0.30-m length of 1-cm-diameter copper tubing, first allowing the water sample to flow from the Kemmerer sampler through the tubing. Once gas bubbles were purged from the tubing, it was crimped on both ends and sealed with refrigerator clamps.

**Helium measurement methods.** Dissolved helium and helium isotope ratios were measured in the USGS Noble Gas Laboratory in Denver, Colorado. The dissolved gases were separated from the formation water in the laboratory and analysed in an ultrahigh-vacuum extraction system. The total pressure of the extracted gases was measured on a capacitance manometer. Helium was separated from the sample gas using a cryogenic cold trap and analysed for isotopic composition statistically on a MAP-215-50 mass spectrometer<sup>38</sup>.

**Helium data interpretation.** Helium accumulation rates were not estimated or calibrated in this study, but were calibrated against  $^{14}\text{C}$  and  $^{36}\text{Cl}$  abundances in another recent study in the near vicinity<sup>25</sup>. The calibration results from that recent study allowed us to make an estimate of the age using helium. We acknowledge that the estimate is a large extrapolation from the measured ages and thus represents an estimate that is far from precise (our error bars are 80 Myr). The validity of the extrapolation is supported, however, by the fact that the helium from the calibration study was extracted and measured from a section of the Potomac Formation very similar to that from which the helium in this study was extracted. In this respect, 'deep crustal' contributions of helium that might be present would have contributed to the calibrated and extrapolated He values in a similar fashion. As additional evidence, deep sources of helium often have a mantle ( $^3\text{He}$ ) signature. We measured  $^3\text{He}/^4\text{He}$  ratios ( $R/R_a$ ) (reported in the main text of this article) and found the value at the deep drill site to be close to zero, suggesting that the helium was solely from crustal *in situ* decay, with a negligible mantle source. In contrast, the helium at the centre of the crater had a value of  $R/R_a$  substantially

above zero, perhaps influenced by a deeper source originating through deep crustal fractures that were generated during the impact. In general, we consider the helium-based age to be one piece of circumstantial evidence in this study. It is not proof of the age, but it is consistent with the other pieces of evidence. The age is further constrained to be not older than Early Cretaceous (by the age of the sediment) and not younger (because the North Atlantic then became hydrologically open).

**Strontium and boron isotopes.** Rock–water interaction is indicated by observing a combination of the dissolved strontium and boron isotopes (Extended Data Fig. 2). The values do not represent modern sea water—especially not boron ( $^{11}\text{B}$ ), which has a modern seawater value of approximately +40 per mil relative to NIST SRM 951 boric acid. The data do trend, however, in the direction of water–rock interaction with igneous or igneous-derived sediments, both of which have typical values that could easily fall in the range indicated by 'Pre-impact basement igneous rock' in Extended Data Fig. 2. Mixing with modern sea water is also suggested by the isotopes, which is consistent with the 35 million years of diffusion that has occurred between the ECNA and modern sea water in the crater.

**Concentration profile simulation.** The process of molecular diffusion and upward advection of groundwater was simulated by solving the advection–diffusion equation for the flow of groundwater in a porous medium<sup>23</sup>:

$$D_m \frac{\partial^2 C}{\partial z^2} - q \frac{\partial C}{\partial z} = \phi \frac{\partial C}{\partial t} \quad (1)$$

where  $D_m$  is the coefficient of molecular diffusion in dimensional units of [ $\text{L}^2/\text{T}$ ];  $C$  is the concentration in [ $\text{M}/\text{L}^3$ ];  $z$  is the vertical distance in [ $\text{L}$ ],  $q$  is the specific discharge in [ $\text{L}/\text{T}$ ],  $\phi$  is porosity, and  $L$ ,  $T$  and  $M$  are the dimensions of length, mass and time, respectively. For the simulation of  $\delta^{37}\text{Cl}$  profiles (see Supplementary Information), equation (1) must be solved for both  $^{35}\text{Cl}$  and  $^{37}\text{Cl}$  and the  $\delta^{37}\text{Cl}$  values calculated from the two results<sup>39</sup>.

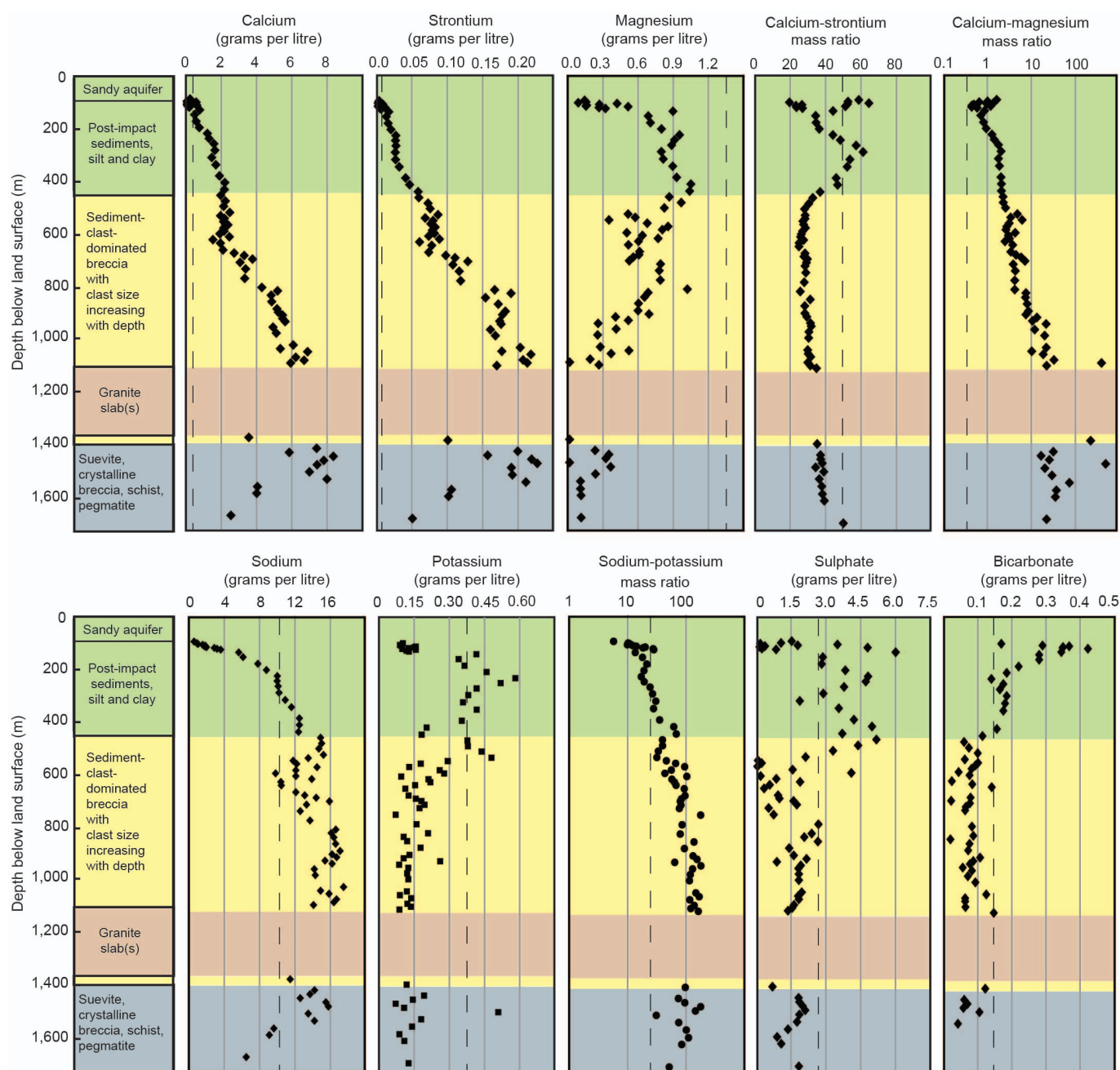
Equation (1) was solved for the vertical concentration profiles of chloride and  $\delta^{18}\text{O}$  in the crater from the land surface down to a depth of 1,400 m (Extended Data Fig. 3). A finite difference approximation method was used with a spatial discretization of 15 m, and time steps of 25,000 years. The finite difference equation was solved in space and time explicitly using an Excel spreadsheet. The upper boundary condition represented the top of the sediment column as it accumulated over 35 million years since impact<sup>40</sup>. The chloride and  $\delta^{18}\text{O}$  boundary conditions at the top of the profile were allowed to switch back and forth between freshwater and seawater conditions, depending on the environment at the time. The upward advective fluid flux  $q$  represents the upward movement of pore water being slowly expelled by compaction of the entire sediment column over time. In reality, the fluid flux would have varied with depth and over time, but here a single value was used as a simplifying assumption. This assumption was adequate to fit the data, and the additional complexity of variations in  $q$  was not justified here, given the scatter in the data that were being fitted. The evolving concentration profiles are shown in Extended Data Fig. 3. Porosity values were assigned that approximated those from the cores<sup>19</sup>—0.45 from 0 m to 230 m, 0.60 from 225 m to 345 m, 0.45 from 345 m to 445 m and 0.3 from 445 m to 1,400 m. As with the fluid flux, the porosity varied more with depth and over time, but this approximation was adequate to fit the data.

The molecular diffusion coefficient and fluid flux value  $q$  were adjusted until a best fit was made with both profiles. The best-fit values and profiles for chloride and  $\delta^{18}\text{O}$  are shown in Extended Data Fig. 5a, e, respectively. The uncertainty in these values was investigated by varying the boundary conditions and parameter values (Extended Data Fig. 5). Simulations without the change in the concentration and sedimentation history at the upper boundary, or either, show poorer fits. The plots shown include adjusting the values of  $D_m$  and  $q$  to obtain the best fit possible. The values of  $D_m$  and  $q$  were also adjusted upward and downward by factors of two, to show the sensitivity of the fit to the parameter. The sum of squared errors is listed for all simulations, and increases substantially for all of the alternative simulation conditions other than the global best-fit values.

- Rostad, C. & Sanford, W. Polar organic compounds in pore waters of the Chesapeake Bay crater deep core hole: character of the dissolved organic carbon and drilling fluid detection. *Geol. Soc. Am. Spec. Pap.* **458**, 891–903 (2009).
- Revesz, K. & Coplen, T. Determination of the  $\delta(^2\text{H}/^1\text{H})$  of water. USGS Techniques and Methods Report, Vol. 10, Ch. C1 (US Geological Survey, 2008).
- Revesz, K. & Coplen, T. Determination of the  $\delta(^{18}\text{O}/^{16}\text{O})$  of water. USGS Techniques and Methods Report, Vol. 10, Ch. C2 (US Geological Survey, 2008).
- Carmody, R. *et al.* Methods for collection of dissolved sulfate and sulfide and analysis of their sulfur isotopic composition. *US Geol. Surv. Open-file Rep.* Report 97–234 (1997).
- Brenna, J. *et al.* High-precision continuous-flow isotope ratio mass spectrometry. *Mass Spectrom. Rev.* **16**, 227–258 (1997).
- Barth, S. Boron isotopic analysis of natural fresh and saline water by negative thermal ionization mass spectrometry. *Chem. Geol.* **143**, 255–261 (1997).

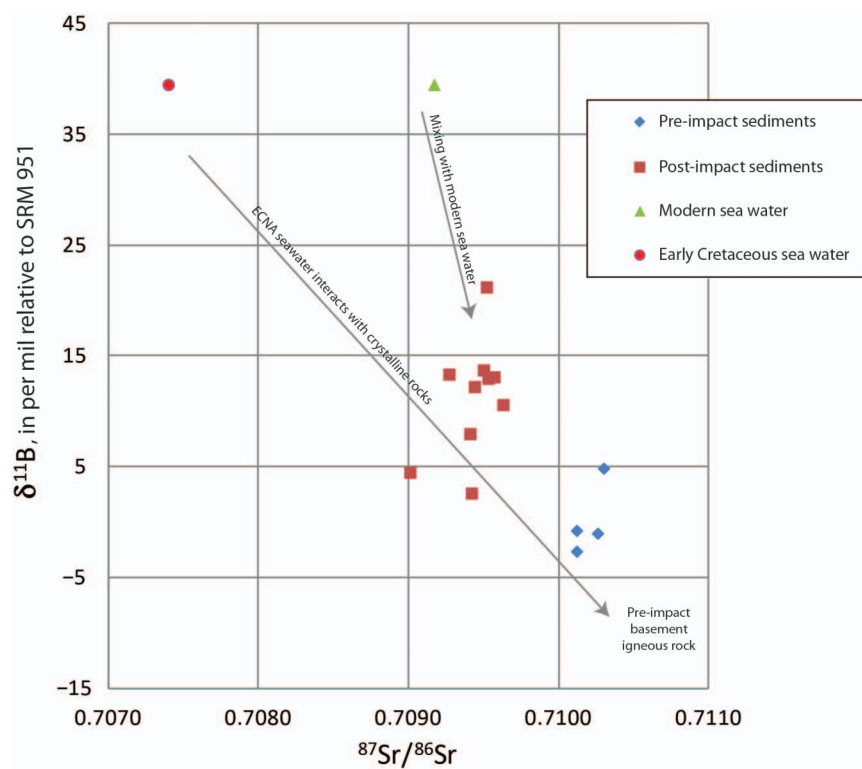
37. Sturchio, N. *et al.* in *Perchlorate, Environmental Occurrence, Interactions and Treatment* (eds Gu, B. & Coates, J. D.) Ch. 5 93–107 (Springer, 2006).
38. Beyerle, U. *et al.* A mass spectrometric system for the analysis of noble gases and tritium from water samples. *Environ. Sci. Technol.* **34**, 2042–2050 (2000).
39. Mazurek, M. *et al.* Natural tracer profiles across argillaceous formations. *Appl. Geochem.* **26**, 1035–1064 (2011).
40. Browning, J. *et al.* Integrated sequence stratigraphy of the postimpact sediments from the Eyreville core holes, Chesapeake Bay impact structure inner basin. *Geol. Soc. Am. Spec. Pap.* **458**, 775–810 (2009).





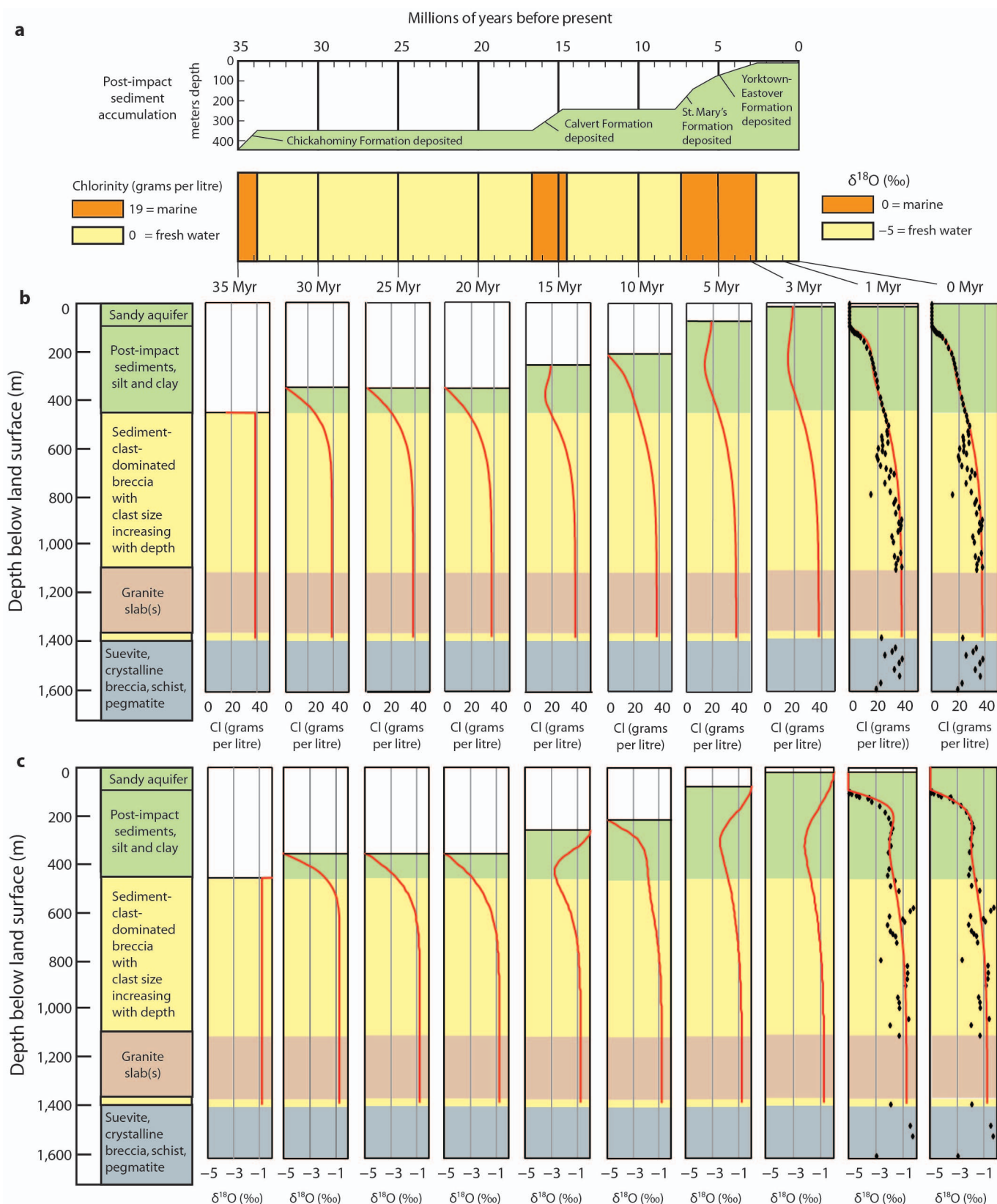
**Extended Data Figure 1 | Results from major cation and anion analyses.** Results from the analyses of major cation and anion concentrations and their ratios in the crater corehole pore waters<sup>19</sup>. Chloride and bromide results are

given in Fig. 2. The vertical dashed black lines represent concentrations or values for modern sea water. For additional discussion see Supplementary Information.



**Extended Data Figure 2 | Plot of strontium and boron isotope measurements.** Plot showing the water–rock interaction of pre-impact igneous basement and resurge blocks according to the mixing lines of dissolved strontium and boron isotopes. Blue diamonds and red squares represent

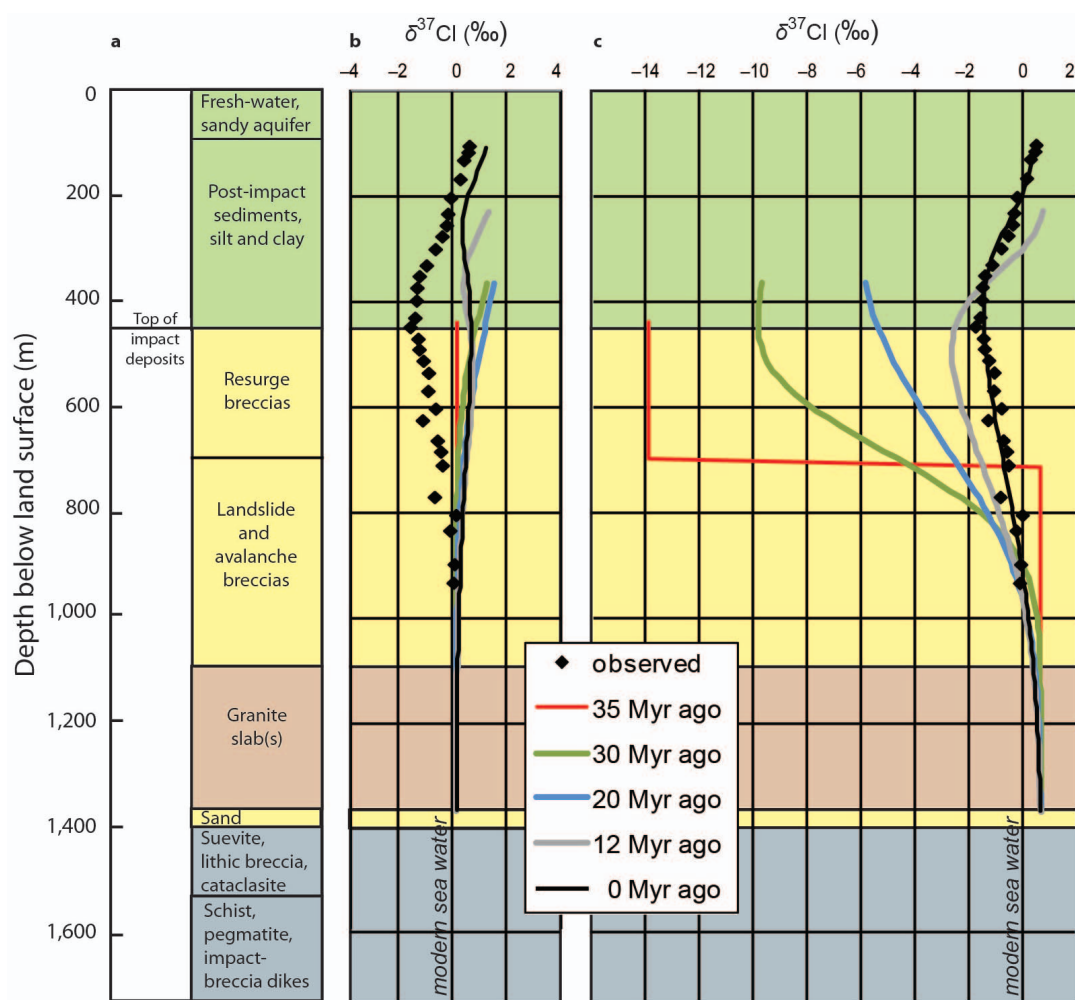
analyses of pore waters from cores of pre-impact and post-impact sediments from the Eyreville drill hole, respectively. For additional discussion see Methods.



**Extended Data Figure 3 | Detailed results from chloride and  $^{18}\text{O}$  transport simulations.** Results from the numerical simulation of chloride and  $\delta^{18}\text{O}$  diffusion and vertical advection. Shown are the transient change in the top model boundary condition over time, including the upward migration of the sediment–water interface and the changes in chlorinity (a), the simulated change in chloride concentration with depth and time (b), and the simulated

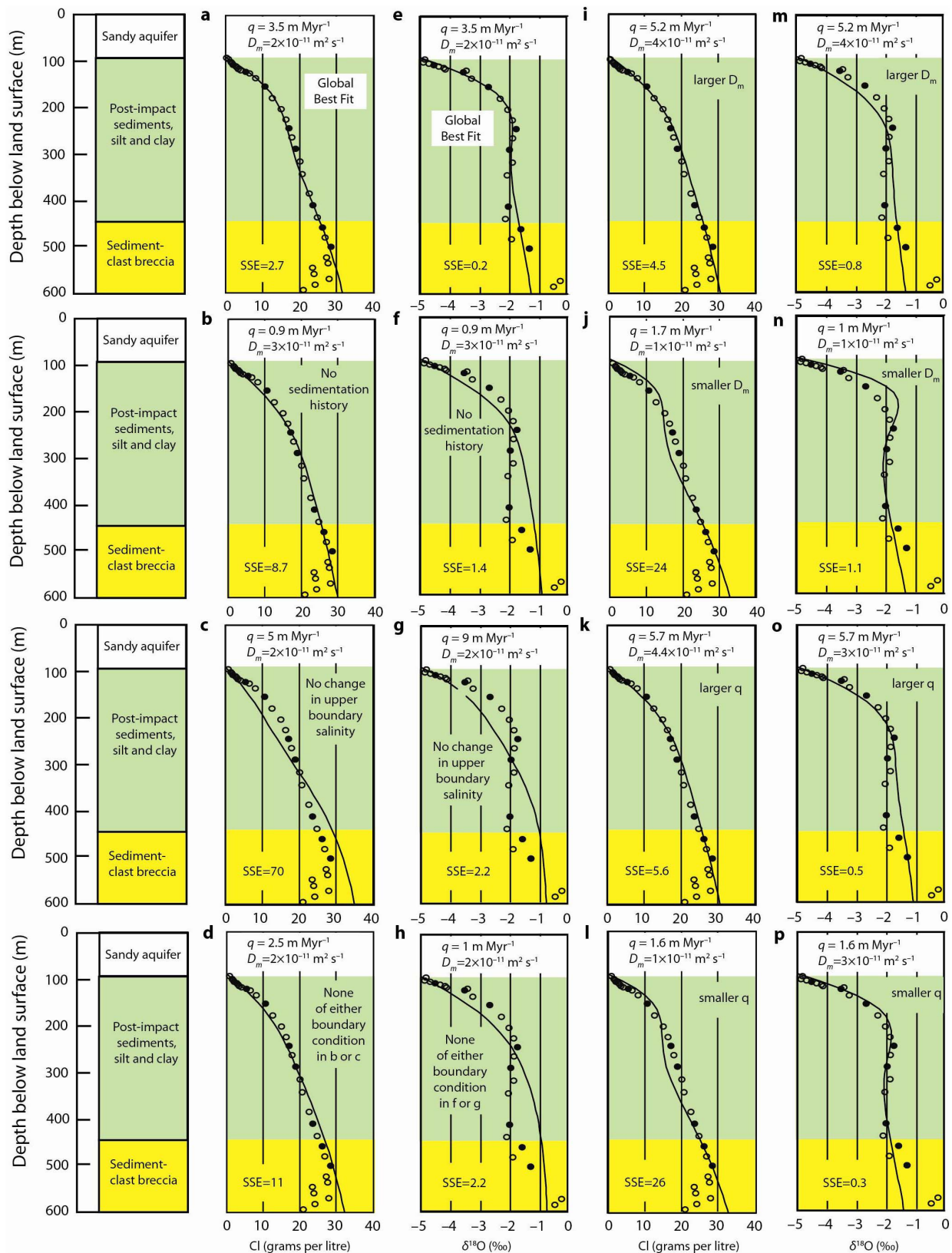
change in  $\delta^{18}\text{O}$  abundances relative to VSMOW (Vienna Standard Mean Ocean Water) with depth and time (c). Red lines are simulated values, and black diamonds shown in the 1-Myr and 0-Myr panels are observation data from cores for comparison. Simulations use the best-fit parameter values of  $D_m = 2 \times 10^{-12} \text{ m}^2 \text{ s}^{-1}$  and  $q = 3.5 \text{ m Myr}^{-1}$ . For additional discussion see Methods.





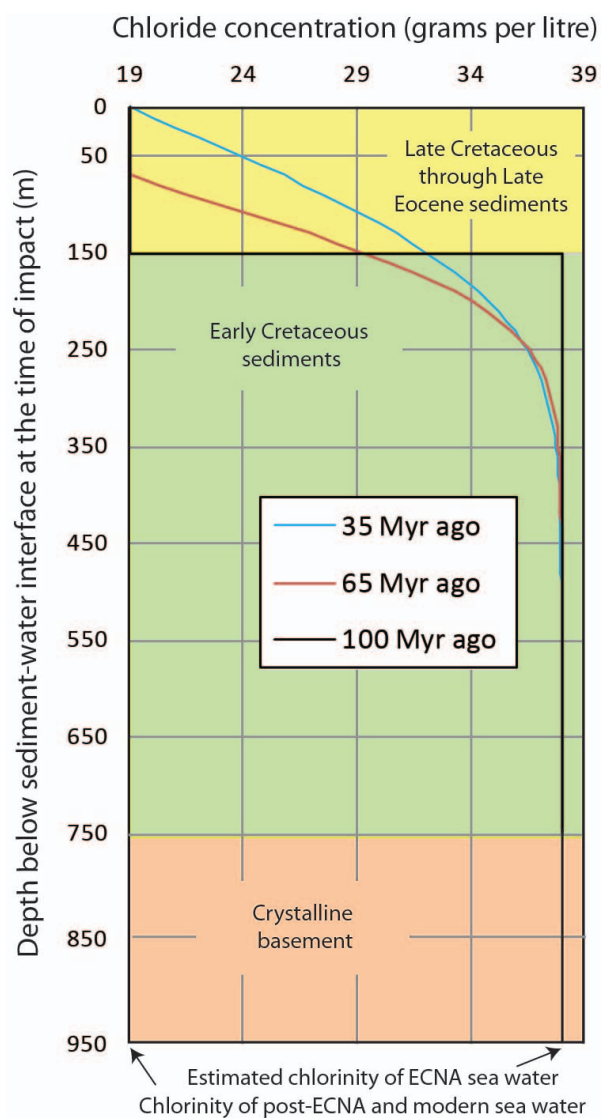
**Extended Data Figure 4 | Stable chloride isotope results.** **a**, Major lithologies drilled at the Chesapeake Bay impact crater, along with observed  $\delta^{37}\text{Cl}$  data and results of numerical simulations of  $\delta^{37}\text{Cl}$  relative to SMOC (standard mean ocean chloride) using the best-fit parameters of the chloride and  $\delta^{18}\text{O}$  advection-diffusion simulations (Fig. 2). Two conceptual models were investigated to explain the observed data, including isotopic fractionation using

a ratio of diffusion coefficients of  $^{37}\text{Cl}/^{35}\text{Cl}$  of 1.002, as observed in other argillaceous materials<sup>39</sup> (**b**) and a  $\delta^{37}\text{Cl}$  anomaly in the resurge breccias created at the time of impact 35 Myr ago (**c**). The value of -14 per mil for the anomaly was the optimal value to fit the observed data. For additional discussion, see Supplementary Information.



**Extended Data Figure 5 | Uncertainty analysis for the chloride and  $^{18}\text{O}$  transport simulations.** Plots showing model uncertainty relative to the most realistic boundary conditions and global best-fit parameter values for Cl (a) and  $\delta^{18}\text{O}$  (e). Alternative boundary conditions include no sedimentation history (b, f), no transient change in the upper boundary salinity (c, g), and neither of these (d, h) for both chloride (b–d) and  $\delta^{18}\text{O}$  (f–h). Variations for the best-fit parameters including half (i, k, m, o) and twice (j, l, n, p) the best-fit values for

both the coefficient of molecular diffusion ( $D_m$ ) (i–l) and the magnitude of the upward fluid flux  $q$  (m–p). Filled and open circles represent those data points used in the sum-of-squared-errors (SSE) calculation, and those not used, respectively. Solid lines are simulation results. For plots i–p one of the two parameters is reset, and the other readjusted to a new best fit. For additional discussion see Methods.



**Extended Data Figure 6 | Simulation of pre-impact chloride profile.**

Simulated chloride concentrations through the pre-impact coastal plain sediments. The black, red and blue lines indicate chloride concentrations at the end of the Early Cretaceous, at the end of the Cretaceous Period and at the time of the bolide impact, respectively. For additional discussion see Supplementary Information.



# The earliest known holometabolous insects

André Nel<sup>1</sup>, Patrick Roques<sup>1</sup>, Patricia Nel<sup>1,2</sup>, Alexander A. Prokin<sup>3</sup>, Thierry Bourgoin<sup>1</sup>, Jakub Prokop<sup>4</sup>, Jacek Szewdo<sup>5</sup>, Dany Azar<sup>6,1</sup>, Laure Desutter-Grandcolas<sup>1</sup>, Torsten Wappler<sup>7</sup>, Romain Garrouste<sup>1</sup>, David Coty<sup>1</sup>, Diying Huang<sup>8</sup>, Michael S. Engel<sup>9,10</sup> & Alexander G. Kirejtshuk<sup>11,1</sup>

The Eumetabola (Endopterygota (also known as Holometabola) plus Paraneoptera)<sup>1</sup> have the highest number of species of any clade, and greatly contribute to animal species biodiversity<sup>2,3</sup>. The palaeoecological circumstances that favoured their emergence and success remain an intriguing question<sup>3–6</sup>. Recent molecular phylogenetic analyses have suggested a wide range of dates for the initial appearance of the Holometabola, from the Middle Devonian epoch (391 million years (Myr) ago) to the Late Pennsylvanian epoch (311 Myr ago<sup>7–12</sup>), and Hemiptera (310 Myr ago<sup>13</sup>). Palaeoenvironments greatly changed over these periods, with global cooling and increasing complexity of green forests<sup>14</sup>. The Pennsylvanian-period crown-eumetabolan fossil record remains notably incomplete<sup>15–19</sup>, particularly as several fossils have been erroneously considered to be stem Holometabola<sup>1,15,20,21</sup> (Supplementary Information); the earliest definitive beetles are from the start of the Permian period<sup>21,22</sup>. The emergence of the hymenopterids, sister group to other Holometabola, is dated between 350 and 309 Myr ago<sup>8,9,12</sup>, incongruent with their current earliest record (Middle Triassic epoch)<sup>1,20</sup>. Here we describe five fossils—a Gzhelian-age stem coleopterid, a holometabolous larva of uncertain ordinal affinity, a stem hymenopterid, and early Hemiptera and Psocodea, all from the Moscovian age—and reveal a notable penecontemporaneous breadth of early eumetabolan insects. These discoveries are more congruent with current hypotheses of clade divergence. Eumetabola experienced episodes of diversification during the Bashkirian–Moscovian and the Kasimovian–Gzhelian ages. This cladogenetic activity is perhaps related to notable episodes of drying resulting from glaciations, leading to the eventual demise in Euramerica of coal-swamp ecosystems, evidenced by floral turnover during this interval<sup>23,24</sup>. These ancient species were of very small size, living in the shadow of Palaeozoic-era ‘giant’ insects. Although these discoveries reveal unexpected Pennsylvanian eumetabolan diversity, the lineage radiated more successfully only after the mass extinctions at the end of the Permian period, giving rise to the familiar crown groups of their respective clades.

Newly discovered species permit a new inventory of the Pennsylvanian Eumetabola. These small insects are scarcely visible against the dark matrix, explaining why they were hitherto generally overlooked. The lack of fossil insects between the Late Devonian and the early Serpukhovian age (Hexapoda gap<sup>25</sup>) could be due in part to a failure to look for such diminutive remains, with most researchers hunting for the more familiar ‘giants’ of these periods. Informative descriptions and details on affinities of the new taxa are given in the Supplementary Information and Methods.

Clade Holometabola  
Stem group to the hymenopterids  
Avioxyelidae fam. nov.  
*Avioxyela gallica* gen. et sp. nov.

**Referred material.** Holotype Avion no. 3 (Avion, Pas-de-Calais, France, Moscovian) (Supplementary Information).

**Etymology.** The genus name is a combination of the type locality Avion and the ‘primitive’ wasp genus *Xyela*. The name is feminine. The specific epithet refers to the ancient Roman name for France. *Avioxyela* is type genus for Avioxyelidae and *A. gallica* is type species for *Avioxyela*.

**Diagnosis.** Forewing characters only. Subcosta posterior zigzagged and short, with cross-veins between it and the cubitus and radius; a very large ‘pterostigmal’ cell between branches of radius anterior (autapomorphy); radius anterior 1 (RA1) and RA2 perpendicular to RA3 and RA4, and to anterior wing margin; radius posterior emerging from radius posterior and median posterior at right angle; radius posterior, median anterior and median posterior simple more or less parallel, with few large rectangular cells between them; a very small triangular cell just distal of the pterostigmal cell and a triangular radial cell near wing apex (synapomorphies with the hymenopterid lineage)<sup>26</sup>; cubitus anterior long, independent of median vein, long cubitus posterior with several posterior branches and one cell between it and the posterior wing margin (Fig. 1a and Supplementary Figs 1 and 2).

Stem group to the coleopterids

Stephanastidae fam. nov.

*Stephanastus polinae* gen. et sp. nov.

**Referred material.** Holotype MNHN.A.49011 (Gzhelian, Commentry, Allier, France).

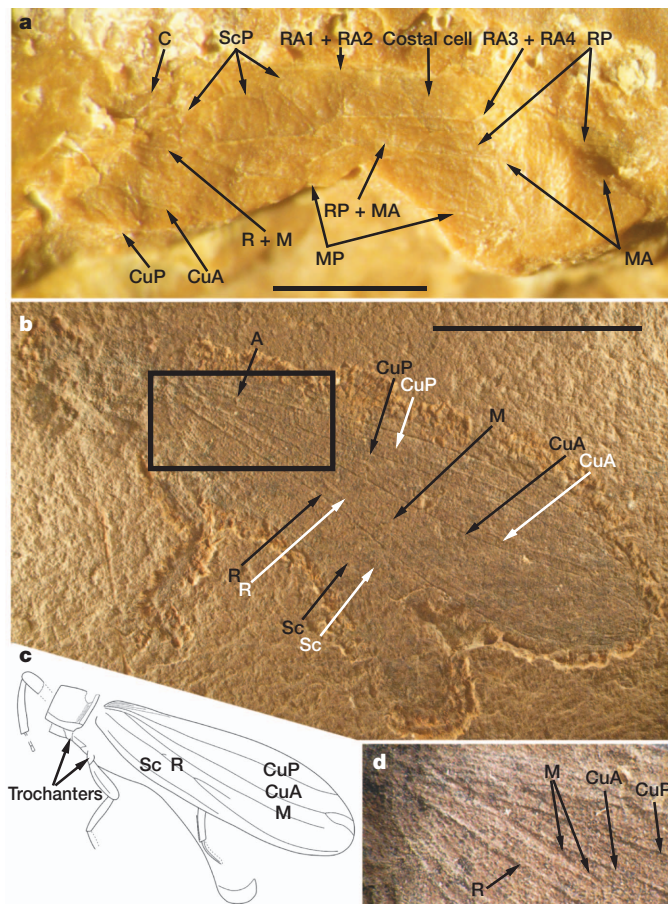
**Etymology.** The genus name is a reference to the Stephanian stage (also known as Gzhelian, originally named after Saint Stephen) and ‘fellow citizen’ in Greek (*‘astos’*: *astos* or *aste* belongs to the *astu* (settlement) and likewise the *polites* or *polititis*). The name is masculine. The species epithet is a matronym honouring Polina G. Kirejtshuk, daughter of one of the authors of the manuscript. *Stephanastus* is type genus for Stephanastidae and *S. polinae* is type species for *Stephanastus*.

**Diagnosis.** Forewing subcosta simple; radius posterior and median simple; concave median and convex cubitus anterior with a common stem separated from radius (synapomorphies with the coleopterids); very long cubitus posterior reaching distal eighth of forewing; anal veins very short at base of forewing (autapomorphies); forewings sclerotized and longer than abdomen, with very fine and dense isodiametric cells forming regular rows generating intercalary veins along anterior edge, more diffuse near posterior margin; no adsutural line (sub-marginal anal vein along the whole posterior edge), medial edge of forewings not forming straight adjoining margins, apices not meeting apically; very large area between cubitus and radius; coxae not projecting, trochanters apparently short; lateral prothoracic carina absent (no prohypomera) (Fig. 1b–d and Supplementary Fig. 3).

Clade Holometabola

*Metabolarva bella* gen. et sp. nov.

<sup>1</sup>CNRS UMR 7205, Muséum National d’Histoire Naturelle, CP 50, Entomologie, 45 Rue Buffon, Paris F-75231, France. <sup>2</sup>AgroParisTech, Département Sciences de la Vie et Santé, 16 rue Claude Bernard, Paris Cedex 05 F-75231, France. <sup>3</sup>Papanin Institute for Biology of Inland Waters, Russian Academy of Sciences, Borok, Nekouzsky District, Yaroslavl Oblast 152742, Russia. <sup>4</sup>Charles University in Prague, Faculty of Science, Department of Zoology, Viničná 7, CZ-128 44, Praha 2, Czech Republic. <sup>5</sup>Department of Palaeozoology, Museum and Institute of Zoology, Polish Academy of Sciences, 64, Wilcza Street, Warszawa PL00-679, Poland. <sup>6</sup>Lebanese University, Faculty of Sciences II, Department of Natural Sciences, Fanar, Fanar – Matn, P.O. Box 26110217, Lebanon. <sup>7</sup>Steinmann-Institut für Geologie, Mineralogie und Paläontologie, Universität Bonn, Nussallee 8, Bonn 53115, Germany. <sup>8</sup>State Key Laboratory of Palaeobiology and Stratigraphy, Nanjing Institute of Geology and Palaeontology, Chinese Academy of Sciences, Nanjing 210008, China. <sup>9</sup>Division of Entomology, Natural History Museum, University of Kansas, Lawrence, Kansas 66045, USA. <sup>10</sup>Department of Ecology & Evolutionary Biology, University of Kansas, Lawrence, Kansas 66045, USA. <sup>11</sup>Zoological Institute of the Russian Academy of Sciences, 1 Universitetskaya Embankment, Saint Petersburg 199034, Russia.



**Figure 1 | Representatives of stem hymenopterids and stem coleopterids.** **a**, *Avioxyela gallica* gen. et sp. nov. imprint. **b–d**, *Stephanastus polinae* gen. et sp. nov. **b**, General habitus (white and black arrows correspond to the veins of the two forewings). **c**, Reconstruction of habitus (numerous intercalary longitudinal pseudo-veins removed for clarity). **d**, Enlargement of the elytra base (black box in **b**). A, anal vein; C, cubitus; CuA cubitus anterior; CuP cubitus posterior; cua-cup, cubital brace; M, median; PC, precosta; R, radius; RA, radius anterior; RP, radius posterior; ScP, subcosta posterior. Scale bars, 1 mm (**a**), 10 mm (**b**).

**Referred material.** Holotype MAS Pal 588 (Moscovian, Piesberg, Lower Saxony, Germany).

**Etymology.** The name of the new genus is formed from the term 'metabolism' and Latin 'larva' (immature active stage of ontogenesis). The name is feminine. The species name *bella* is in reference to the good state of preservation. *Metabolarva bella* is type species of the genus *Metabolarva*.

**Diagnosis.** Larviform animal of eruciform oligopod type; prothorax with paired wide sclerotized plates narrowly separated along midline; abdomen membranous, apparently with eight segments; probable stigmal plate bearing a pair of spiracular orifices; distal preapical abdominal segments bearing in total four pairs of thick appendages forked into two (anterior and posterior) processes (Fig. 2a, b, and Supplementary Fig. 4).

Clade Paraneoptera

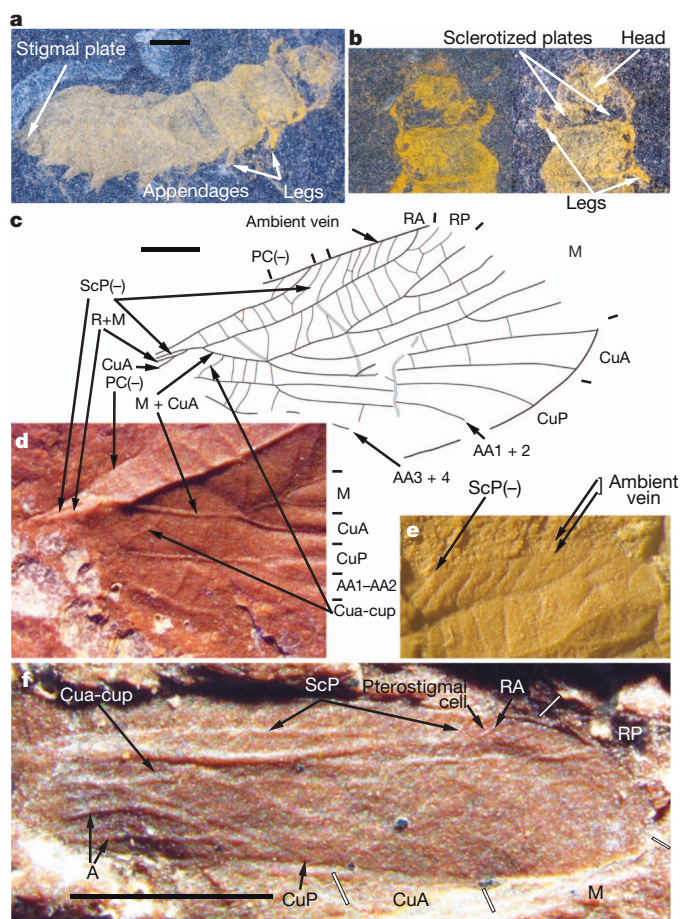
Order Hemiptera

Family Aviorrhynchidae fam. nov.

*Aviorrhyncha magnifica* gen. et sp. nov.

**Referred material.** Holotype Avion no. 2 (Avion, Pas-de-Calais, France, Moscovian) (Supplementary Information).

**Etymology.** The generic name refers to the type locality Avion, and *rrhyncha* refers to the common suffix of related suborders (for example, Sternorrhyncha, Auchenorrhyncha). The name is feminine. The species epithet refers to the wonderful state of preservation of the holotype.



**Figure 2 | Holometabolous larva and representatives of stem Psocodea and Euhemiptera.** **a**, **b**, *Metabolarva bella* gen. et sp. nov. **a**, Photograph. **b**, Reconstruction of habitus. **c–e**, *Aviorrhyncha magnifica* gen. et sp. nov. **c**, Reconstruction of forewing. —, concave veins; AA1 + 2, first anal anterior vein; AA3 + 4, second anal anterior vein. Short lines outside of the wing illustration mark the limits of different veins along the wing margin. **d**, Detail of wing base. **e**, Detail of anterior margin. **f**, *Westphalopsocus pumilio* gen. et sp. nov., forewing. PC, precosta. Scale bars, 1 mm (**a**, **c**, **f**).

*Aviorrhyncha* is type genus for Aviorrhynchidae and *A. magnifica* is type species for *Aviorrhyncha*.

**Diagnosis.** Forewing with a cua-cup vein, long concave precosta, very short concave subcosta posterior fused with radius near wing base, re-emerging distally, a fringed ambient vein (synapomorphies with Paraneoptera or Euhemiptera); a series of veinlets between radius and anterior ambient vein; *areola postica* with long simple branches; simple cubitus posterior and first anterior anal (AA1 + 2) fused near wing base and separating again distally (Fig. 2c–e and Supplementary Figs 5 and 6a).

Order Psocodea

Family Westphalopsocidae fam. nov.

*Westphalopsocus pumilio* gen. et sp. nov.

**Referred material.** Holotype Avion no. 4 (Avion, Pas-de-Calais, France, Moscovian) (Supplementary Information).

**Etymology.** The generic name refers to the Westphalian stage and *Psocus*, common stem of many generic names in the Psocodea. The name is masculine. The species epithet refers to the very small size of the species. *Westphalopsocus* is type genus for Westphalopsocidae and *W. pumilio* is type species for *Westphalopsocus*.

**Diagnosis.** Forewing narrow elongate with parallel margins, with a cua-cup cross-vein; two anal veins fused apically; cubitus anterior forming an *areola postica*; subcosta posterior distally fused with radius anterior and re-emerging near its apex; a pterostigmal cell (characters of Psocodea)



between apices of subcosta posterior and radius anterior; a series of cross-veins between subcosta posterior and cubitus (autapomorphy); anterior branch of radius posterior emitting four short anterior branches; three posterior median branches; cubitus posterior simple; two anal veins coalescing and fusing apically for a distance (Fig. 2f and Supplementary Fig. 7).

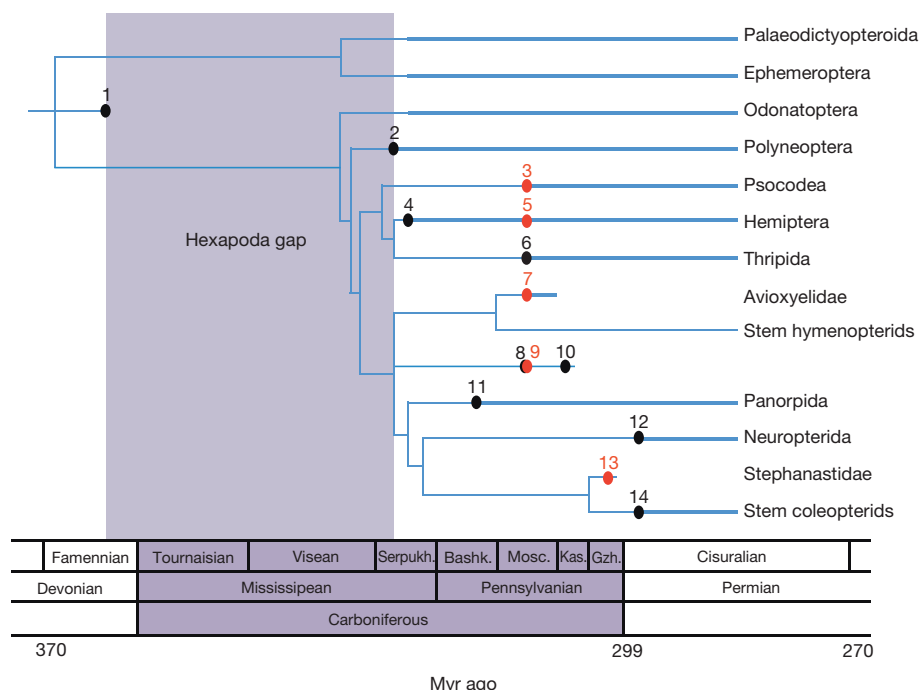
These taxa alter current views on insect diversity during the Pennsylvanian, a critical period in hexapod diversification (Fig. 3 and Table 1). *Avioxyela gallica* is the earliest and first Carboniferous species that can be assigned to the stem lineage of the hymenopterids, one of the three main ecologically ubiquitous insect clades. The approximately 90-Myr gap between this fossil and the earliest crown hymenopterid remains incongruent with the molecular dating of all principal hymenopterid clades from late in the Palaeozoic and early in the Mesozoic era<sup>12</sup>. Molecular studies have failed to account for some Paleozoic ‘mecopteroids’, which are probably more closely allied to *Avioxyela* and the hymenopterids. The previous discovery of the Bashkirian panorpidan *Westphalomerope* indicated that the stem Neuropterodea (the coleopterids plus Neuropterida) is of coeval age<sup>16</sup>. True Neuropterida and definitive tsekaridocoleid beetles are known in the lowermost Permian<sup>20,22</sup>. The Gzhelian *Stephanastus* is the oldest known representative of the stem group to the coleopterids (Coleoptera plus Strepsiptera), demonstrating that the Neuropterida are at least latest Pennsylvanian in age. The crown group Hymenoptera (common ancestor of Xyelidae and Neohymenoptera) perhaps originated during the Permian–Triassic from a grade of stem hymenopterids such as *Avioxyela*. Earlier alleged stem groups to the Holometabola are actually Polyneoptera (Supplementary Information), a group that is already well represented in the Pennsylvanian. The key innovation of the holometabolous larva was probably among the factors that purportedly fuelled their diversification. Such larvae must have already been present during the Pennsylvanian as evidenced by the presence of definitive holometabolans described herein. Given that the putative Moscovian larva *Srokalarva bertei* remains to be formally described<sup>18</sup> (see Supplementary Information), the Moscovian *Metabolarva*

presented here is the oldest direct record of a holometabolous larva. It is likely to have had a metapneustic respiratory system with spiracular orifices at the anal end of the body, and to have lived in a semi-aquatic or moist environment. This eruciform larva is in accordance with the idea that this larval type may represent the groundplan for Holometabola, rather than the campodeiform type proposed earlier<sup>20</sup>.

Pennsylvanian Paraneoptera are currently represented by the Serpukhovian hemipteran *Protoprosbole*, the Moscovian thripidan *Westphalothripides*<sup>17</sup>, and the stem paraneopteran *Bruayaphis*<sup>15</sup>. The Moscovian *Westphalopsocus* therefore represents the oldest evidence of Psocodea, extending the lineage back from earlier records based on the Permian *Zygopsocus*<sup>15</sup>. The Moscovian *Aviorrhyncha* is the second Pennsylvanian hemipteran. Its proposed attribution to the Euhemiptera implies a separation from the Sternorrhyncha dating back at least to the Moscovian, older than suggested by molecular dating<sup>13</sup>.

The new fossil material highlights the diversity of Pennsylvanian Eumetabola, involving stem groups to the major lineages Hymenoptera, Panorpida, Neuropterodea, Thripida, Hemiptera and Psocodea. These discoveries are congruent with recent attempts to estimate divergence times based on sequence data, although those studies have widely failed to distinguish between the divergences of the stem groups, the crown groups of which may have appeared at different times and much later. These data are congruent with earlier evidence suggesting that the initial phases of the diversification of principal insect lineages, is much older than previously surmised<sup>1,25</sup>. Taking into account the presence of taxa such as *Rhyniognatha hirsti* in the Pragian<sup>1</sup>, the data seem to indicate that insects are likely to have originated around the Silurian–Devonian boundary and diversified between the Devonian and the early Pennsylvanian. The data suggest that the foundations of the eventually hyperdiverse Holometabola, comprising most modern-day insect species, were already well established in the Pennsylvanian.

The oldest Permian beetles and the earliest Devonian Pterygota were very small, considerably smaller than many of their descendants<sup>22,25</sup>. Except for the Gzhelian *Stephanastus*, all of the Pennsylvanian adult



**Figure 3 | Phylogeny of main insect clades.** *Strudiella devonica* Garrouste *et al.*, 2012 (ref. 25) (1); wing of an archaeorthopteran, Prokop *et al.*, 2005 (2); *Westphalopsocus pumilio* gen. et sp. nov. (3); *Protoprosbole straeleni* Laurentiaux, 1952 (4); *Aviorrhyncha magnifica* gen. et sp. nov. (5); *Westphalothripides oudardi* Nel *et al.*, 2012 (ref. 15) (6); *Avioxyela gallica* gen. et sp. nov. (7); *Metabolarva bella* gen. et sp. nov. (8); *Srokalarva bertei*

Kukulová-Peck, 1997 (9); *Pteridotorichnos stipitopteri* Labandeira and Phillips, 1996 (10); *Westphalomerope maryvonneae* Nel *et al.*, 2007 (ref. 16) (11); undescribed Neuroptera of Obora (Czech Republic) (12); *Stephanastus polinae* gen. et sp. nov. (13); *Coleopsis archaica* Kirejtshuk *et al.*, 2013 (ref. 22) (14). Bashk., Bashkirian; Gzh., Gzhelian; Kas., Kasimovian; Mosc., Moscovian; Serpukh., Serpukhovian.



**Table 1 | Carboniferous Paraneoptera and Holometabola**

Clade	Genus	Fossil	Length	Stratigraphic stage	Locality	Ref.
Hemiptera	<i>Protoprosbole</i>	Wing	14 mm	Serpukhovian	Forchie, Belgium	15
Hemiptera	<i>Aviorrhyncha</i>	Wing	7.5 mm	Moscovian	Avion, France	Present paper
Psocodea	<i>Westphalopsocus</i>	Wing	3.6 mm	Moscovian	Avion, France	Present paper
Paraneoptera	<i>Bruayaphis</i>	Wing	3.8 mm	Moscovian	Bruay-en-Artois, France	15
Thripida	<i>Westphalothripides</i>	Wing	7.4 mm	Moscovian	Avion, France	17
Holometabola (uncertain)	<i>Baryshnyala</i>	Wing	9.6 mm	Bashkirian	Hagen-Vorhalle, Germany	19
Panorpida	<i>Westphalomerope</i>	Wing	4.8 mm	Bashkirian	Bruay-la-Bussière, France	16
Stem hymenopterid	<i>Avioxyela</i>	Wing	5.5 mm	Moscovian	Avion, France	Present paper
Holometabola	<i>Srokalarva</i>	Larva	30 mm	Moscovian	Mazon Creek, USA	17
Holometabola	<i>Metabolarva</i>	Larva	9.1 mm	Moscovian	Osnabrück, Germany	Present paper
Holometabola	<i>Pteridotorichnos</i>	Gall	5 mm	Kasimovian	Illinois basin, USA	18
Stem coleopterid	<i>Stephanastus</i>	Body with wings	25.5 mm (wing)	Gzhelian	Commentry, France	Present paper

Eumetabola were also distinctively miniaturized, with wing lengths ranging between 3.8 mm and 14.0 mm, range 10.2 mm, in contrast to 120 mm for Recent representatives (see Fig. 4). It would seem that these clades followed Cope's rule<sup>27,28</sup>; that is, the tendency for organisms in certain lineages to increase in size over time. In addition, the Serpukhovian Eugeopteridae, the most basal clade of the Odonatoptera, were of relatively small proportions, with wing lengths ranging between 35 and 45 mm, whereas the 'giant' Meganeuridae (griffinflies) included the largest of all known insects with wing lengths between 50 and 330 mm. These facts contradict the findings in a previous paper<sup>29</sup> in which it was supposed that this rule was not obeyed by insects of the Pennsylvanian. Although modern eumetabolous and non-eumetabolous insects have similarly smaller ranges of body sizes, this was not the case during the Pennsylvanian, as the known non-eumetabolous insects had wing lengths ranging between approximately 10 mm and 330 mm (Fig. 4). The presence of small, non-eumetabolous insects in the Moscovian of Avion (Supplementary Fig. 8) shows that their absence from other localities may be an artefact of biased collecting towards larger, showier hexapods. Conversely, the absence of larger Eumetabola during the Pennsylvanian is assuredly not an artefact. Only a few palaeopteran lineages were affected by gigantism, suggesting that this phenomenon

was not solely due to the Pennsylvanian hyperoxic conditions that should have affected all clades<sup>30</sup>. It could be assumed that a combination of palaeoecological, physiological and environmental factors stimulated the growth of 'giant' insects among certain clades and the miniaturization of the Eumetabola during the time interval.

The 'key innovation' of metamorphosis, present in holometabolous larvae and some Paraneoptera, may have fostered the longevity of these taxa during the Pennsylvanian climatic oscillations<sup>23</sup> and the later Permian–Triassic extinction event<sup>5</sup>. The Holometabola, and the Paraneoptera to a lesser degree, experienced an extraordinary Triassic diversification, after the extinction of several ecologically dominant palaeopteran groups at the end of the Permian. Although holometabolans comprise some of the largest modern insects, their ancient progenitors apparently were very small. Much like the early, tiny mammals that skirted dinosaurian diversity many eons later, the Pennsylvanian Holometabola were dwarfed by but existed alongside the more ubiquitous and famous palaeopteran and polyneopteran insects of that bygone age<sup>18</sup>.

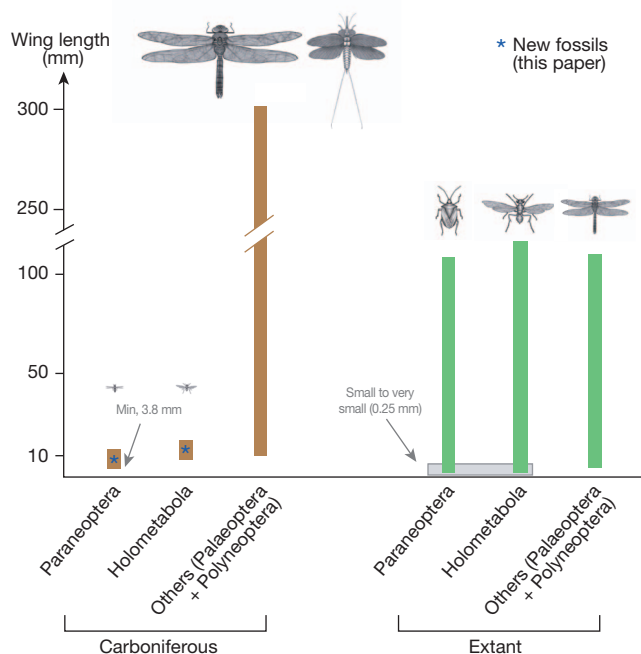
## METHODS SUMMARY

It was necessary to observe each small piece of rock under stereomicroscope to detect the insect fossils. The specimens are provisionally housed at the Muséum National d'Histoire Naturelle, Paris, France, and were prepared using sharp spines. Photographs were taken with a Canon D550 digital camera with reverse lens MP-E 65 mm, with dry specimens, and line drawings were prepared using a camera lucida on an Olympus SZX-9 stereomicroscope. Original photographs were processed using the image-editing software Adobe Photoshop CS. Some specimens were examined with a Tescan Vega LSU scanning electron microscope.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 July; accepted 3 September 2013.

Published online 16 October 2013.



**Figure 4 | Comparison of the sizes of Carboniferous fossil Pterygota and extant Pterygota.** Illustrations of insects demonstrate the different clades (two are given for Palaeoptera and Polyneoptera of the Carboniferous, to show their diversity), their sizes are used to show approximately which are larger than others, but they are not to scale. Smallest extant insects are 0.25 mm long. Smallest carboniferous wings are 3.8 mm long (min. 3.8 mm).

- Grimaldi, D. A. & Engel, M. S. *Evolution of the Insects*. (Cambridge Univ. Press, 2005).
- Wiegmann, B. M. & Kim, J.-W. in *The Timetree of Life* (eds Hedges, S. B. & Kumar, S.) 260–263 (Oxford University Press, 2009).
- Mayhew, P. J. Why are there so many insect species? Perspectives from fossils and phylogenies. *Biol. Rev. Camb. Philos. Soc.* **82**, 425–454 (2007).
- Bernays, E. A. Evolutionary contrasts in insects: nutritional advantages of holometabolous development. *Physiol. Entomol.* **11**, 377–382 (1986).
- Yang, A. S. Modularity, evolvability, and adaptive radiations: a comparison of the hemi- and holometabolous insects. *Evol. Dev.* **3**, 59–72 (2001).
- McPeck, M. A. & Brown, J. M. Clade age and not diversification rate explains species richness among animal taxa. *Am. Nat.* **169**, E97–E106 (2007).
- Rehm, P. et al. Dating the arthropod tree based on large-scale transcriptome data. *Mol. Phylogenet. Evol.* **61**, 880–887 (2011).
- Warnock, R. C. M., Yang, Z. & Donoghue, P. C. J. Exploring uncertainty in the calibration of the molecular clock. *Biol. Lett.* **8**, 156–159 (2012).
- Wiegmann, B. M. et al. Single-copy nuclear genes resolve the phylogeny of the holometabolous insects. *BMC Biol.* **7**, 34 (2009).
- Yeates, D. K., Cameron, S. L. & Trautwein, M. A view from the edge of the forest: recent progress in understanding the relationships of the insect orders. *Aust. J. Entomol.* **51**, 79–87 (2012).
- Wheat, C. W. & Wahlberg, N. Phylogenomic insights into the Cambrian explosion, the colonization of land and the evolution of flight in Arthropoda. *Syst. Biol.* **62**, 93–109 (2013).
- Ronquist, F. et al. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Syst. Biol.* **61**, 973–999 (2012).

13. Song, N. & Liang, A.-P. A preliminary molecular phylogeny of planthoppers (Hemiptera: Fulgoroidea) based on nuclear and mitochondrial DNA sequences. *PLoS ONE* **8**, e58400 (2013).
14. Fiz-Palacios, O., Schneider, H., Heinrichs, J. & Savolainen, V. Diversification of land plants: insights from a family-level phylogenetic analysis. *BMC Evol. Biol.* **11**, 341 (2011).
15. Nel, A. *et al.* Traits and evolution of wing venation pattern in paraneopteran insects. *J. Morphol.* **273**, 480–506 (2012).
16. Nel, A. *et al.* The earliest holometabolous insect: a “crucial” innovation with delayed success (Insecta Protomeropina Protomeropidae). *Ann. Soc. Entomol. Fr. (NS)* **43**, 349–355 (2007).
17. Nel, P. *et al.* From Carboniferous to recent: wing venation enlightens evolution of thysanopteran lineage. *J. Syst. Palaeontology* **10**, 385–399 (2012).
18. Labandeira, C. C. Evidence for an earliest Late Carboniferous divergence time and the early larval ecology and diversification of major Holometabola lineages. *Entomol. Amer.* **117**, 9–21 (2011).
19. Ilger, J.-M. & Brauckmann, C. The smallest Neoptera (Baryshnyalidae fam. n.) from Hagen-Vorhalle (early Late Carboniferous: Namurian B; Germany). *ZooKeys* **130**, 91–102 (2011).
20. Rasnitsyn, A. P. & Quicke, D. L. J. *History of Insects* (Kluwer, 2002).
21. Kukulová-Peck, J. & Beutel, R. G. Is the Carboniferous †*Adiphebia lacoana* really the “oldest beetle”? Critical reassessment and description of a new Permian beetle family. *Eur. J. Entomol.* **109**, 633–645 (2012).
22. Kirejtshuk, A. G. *et al.* Evolution of the elytral venation and structural adaptations in the oldest Palaeozoic beetles (Insecta: Coleoptera: Tshekardocoleidae). *J. Syst. Palaeontology*. (in the press).
23. Davydov, V. I., Korn, D. & Schmitz, M. D. in *The Geologic Time Scale* (eds Gradstein, F., Ogg, J., Schmitz, M. & Ogg, G.) 603–651 (Elsevier, 2012).
24. Phillips, T. L. & Peppers, R. A. Changing patterns of Pennsylvanian coal-swamp vegetation and implications of climatic control on coal occurrence. *Int. J. Coal Geol.* **3**, 205–255 (1984).
25. Garrouste, R. *et al.* A complete insect from the Late Devonian period. *Nature* **488**, 82–85 (2012).
26. Kukulová-Peck, J. & Lawrence, J. F. Relationships among coleopteran suborders and major endoneopteran lineages: evidence from hind wing characters. *Eur. J. Entomol.* **101**, 95–144 (2004).
27. Beckemeyer, R. J. & Hall, J. D. The entomofauna of the Lower Permian fossil insect beds of Kansas and Oklahoma, USA. *Afr. Invertebr.* **48**, 23–39 (2007).
28. Hone, D. W. E. & Benton, M. J. The evolution of large size: how does Cope's Rule work? *Trends Ecol. Evol.* **20**, 4–6 (2005).
29. Chown, S. L. & Gaston, K. J. Body size variation in insects: a macroecological perspective. *Biol. Rev. Camb. Philos. Soc.* **85**, 139–169 (2010).
30. Clapham, M. E. & Karr, J. A. Environmental and biotic controls on the evolutionary history of insect body size. *Proc. Natl Acad. Sci. USA* **109**, 10927–10930 (2012).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank C. C. Labandeira for comments on the first version of the manuscript that helped to improve the paper. We are grateful to A. P. Rasnitsyn, S. I. Golovach and B. R. Striganova, M. Fikáček, A.A. Przhiboro, R. Beutel, T. Hörschemeyer and V. Krassilov for early discussions. C. Garrouste and P. A. Kirejtshuk assisted in the preparation of illustrations for this publication. Financial support was provided by the Grant Agency of the Czech Republic no. P210/10/0633 (to J.P.) and the German Science Foundation WA 1492/6-1 (to T.W.). The study was supported by the program for visiting researchers and professors of the Smithsonian Institution National Museum of Natural History (NMNH) and partly carried out within the framework of the program of the Presidium of the Russian Academy of Sciences ‘Problems of the origin of life and formation of the biosphere’. A.A.P. and A.G.K. were supported by the Russian Foundation of Basic Research (grant 12-04-00663-a). This paper is a participation to the team project ‘Biodiversity: Origin, Structure, Evolution and Geology’ allotted to D.A. by the Lebanese University.

**Author Contributions** A.N., P.R., P.N., A.A.P., T.B., J.P., J.S., D.A., R.G., D.C., D.H., M.S.E. and A.G.K. participated in morphological studies and prepared the manuscript. P.N. and D.C. prepared the figures. P.R. discovered the fossils. The authors of the taxonomic data are associated with the names of the species in the Supplementary Information. A.N. designed the program. M.S.E. and A.G.K. are last authors with equal rank.

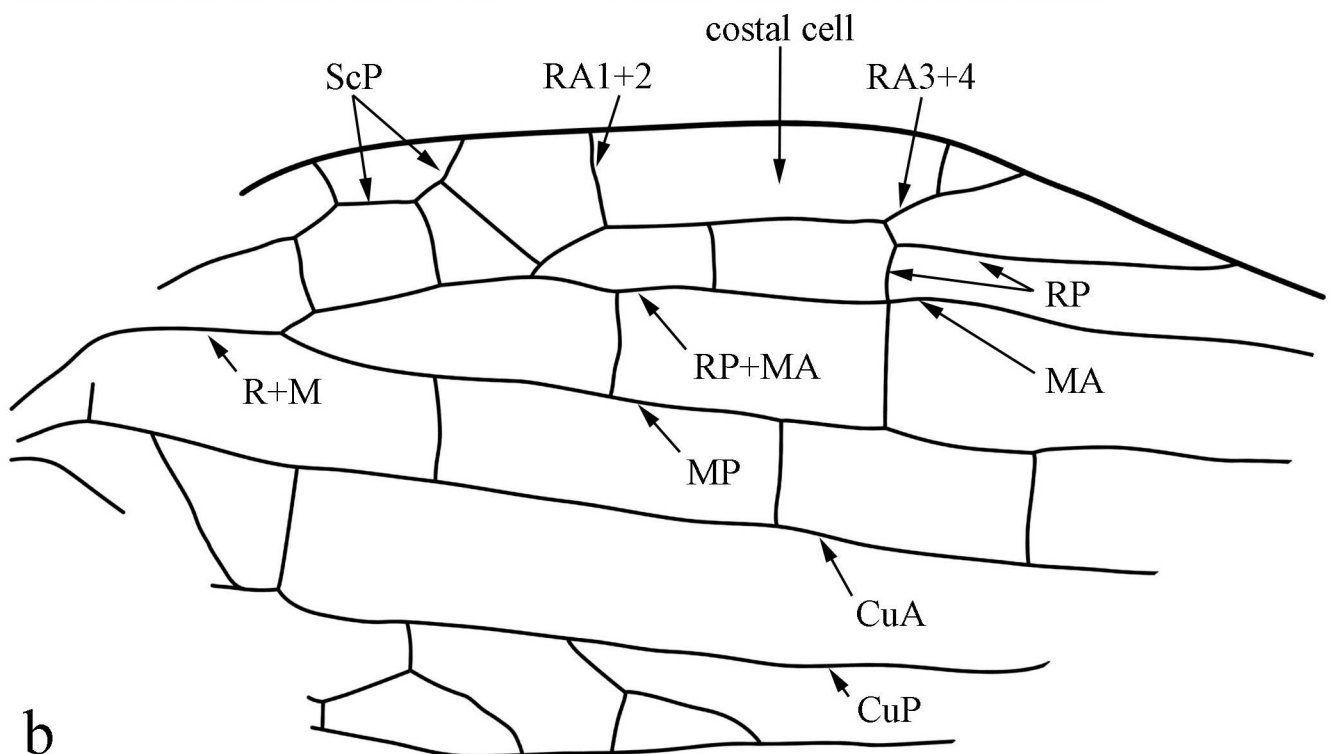
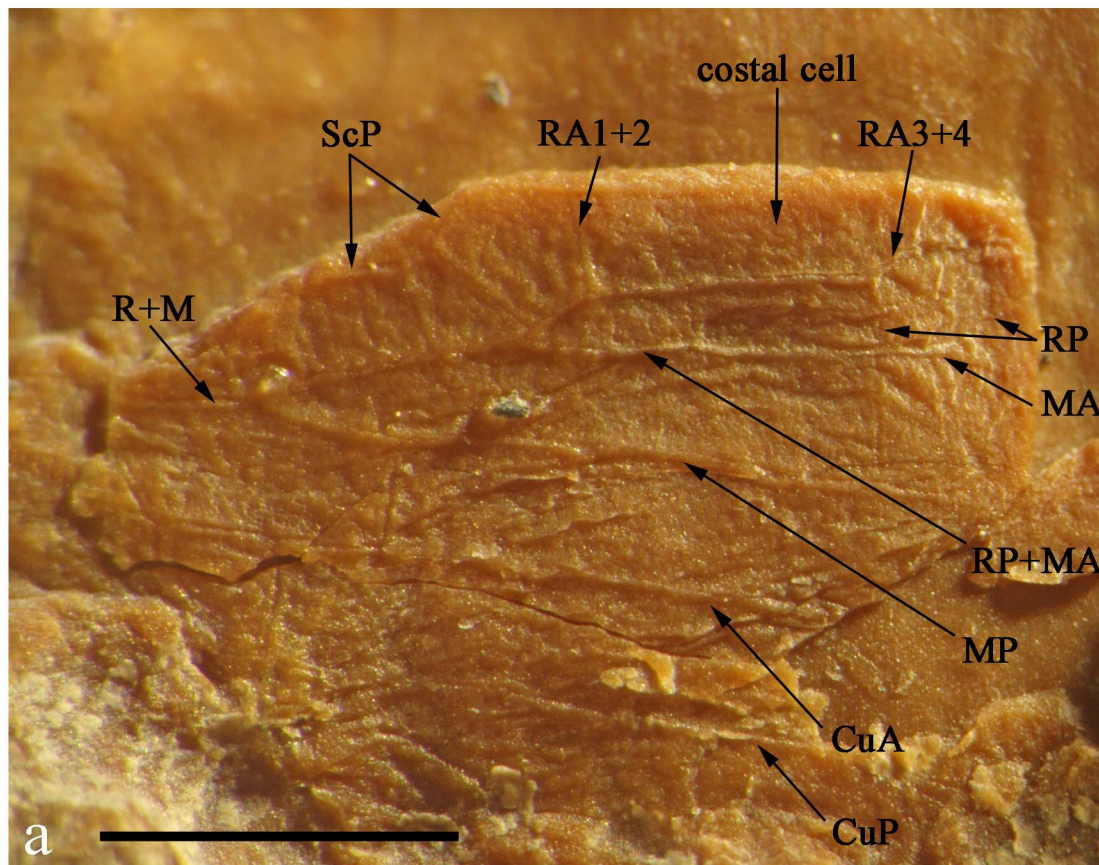
**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.N. (anel@mnhn.fr).

## METHODS

It was necessary to observe each small piece of rock under a stereomicroscope to detect the insect fossils. The specimens are provisionally housed at the Muséum National d'Histoire Naturelle (MNHN), Paris, France, and were prepared using sharp spines. Photographs were taken with a Canon D550 digital camera with reverse lens MP-E 65 mm, with dry specimens, and line drawings were prepared using a camera lucida on an Olympus SZX-9 stereomicroscope. Original photographs were processed using the image-editing software Adobe Photoshop CS. Some specimens were examined with a Tescan Vega LSU scanning electron microscope. Fossils from the Avion outcrop (Moscovian, Department of Pas-de-Calais, France) are in the P. Roques collection, currently stored in the Entomological

Laboratory, MNHN, Paris, and to be deposited in the Musée Géologique Pierre Vetter, Decazeville, France. Standard wing venation nomenclature is followed for Hymenoptera<sup>26</sup>, Coleoptera<sup>22</sup> and Paraneoptera<sup>15</sup>. The diagnoses given in the main text are the same for the familial, generic and specific levels because all are presently monospecific. The Zoobanks LSIDs (Life Science Identifiers) for the new genera and species are as follows: *Avioxyela gallica* LSID, urn:lsid:zoobank.org:act:A9759D8D-785F-472B-A99A-9FE31622CCBC; *Stephanastus polinae* LSID, urn:lsid:zoobank.org:act:A75F53A2-0BE2-425D-97B4-33A572F20D73; *Metabolarva bella* LSID, urn:lsid:zoobank.org:act:7D1BE021-C4C3-4760-B8C7-6099454C2A8F; *Aviorrhyncha magnifica* LSID, urn:lsid:zoobank.org:act:41D4931A-57A9-42FD-8F01-7F909906782E; *Westphalopsocus pumilio* LSID, urn:lsid:zoobank.org:act:68C4E20D-1472-49FD-9826-24CE88A0CACE.

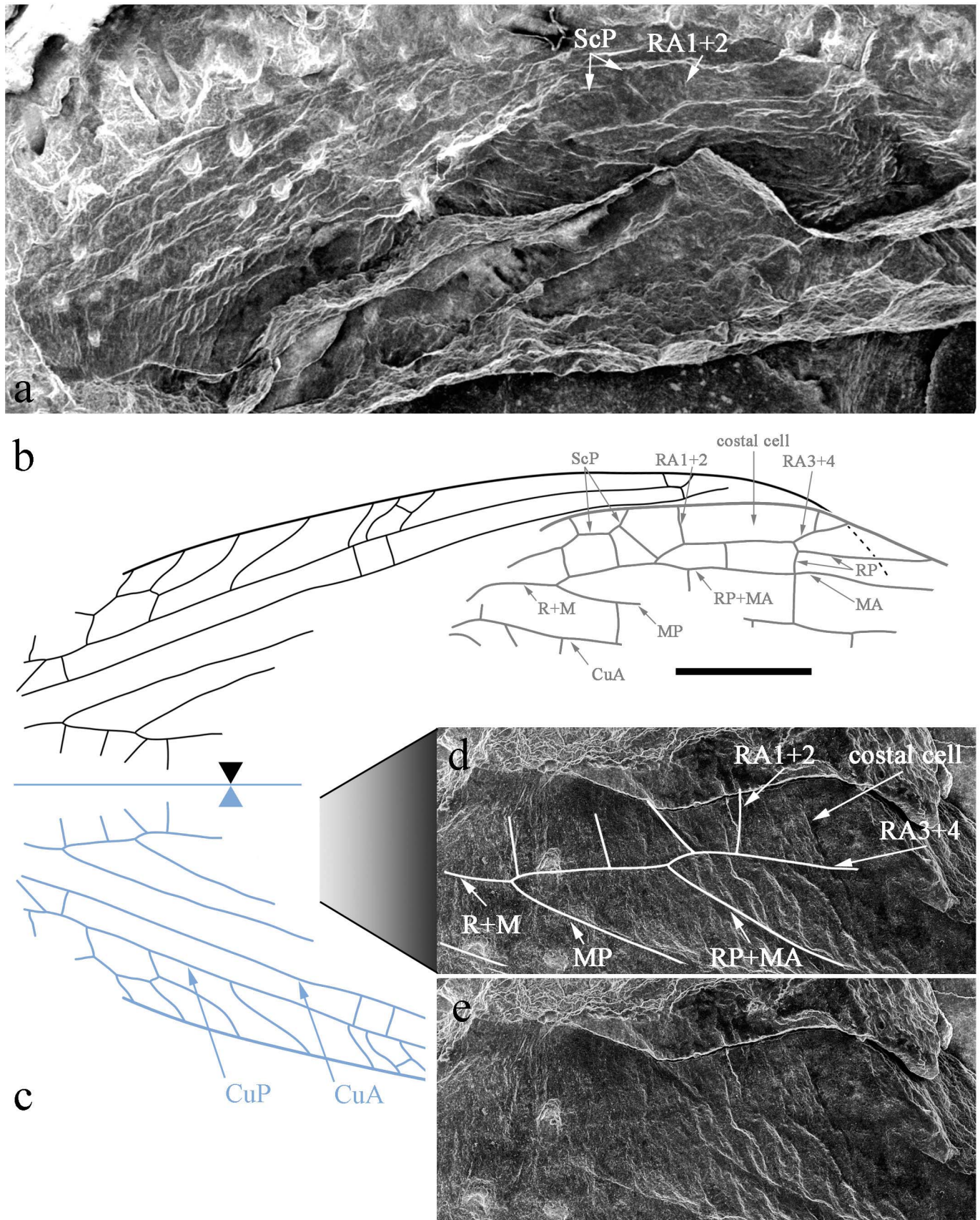




**Extended Data Figure 1 | Representative of the stem hymenopterids, *Avioxyela gallica* gen. et sp. nov.** a, Photograph of counterpart. b, Reconstruction of forewing based on part and counterpart. CuA, cubitus

anterior; CuP, cubitus posterior; MA, median anterior; MP, median posterior; R, radius; RA, radius anterior; RP, radius posterior; ScP, subcosta posterior. Scale bar, 1 mm (a, b).

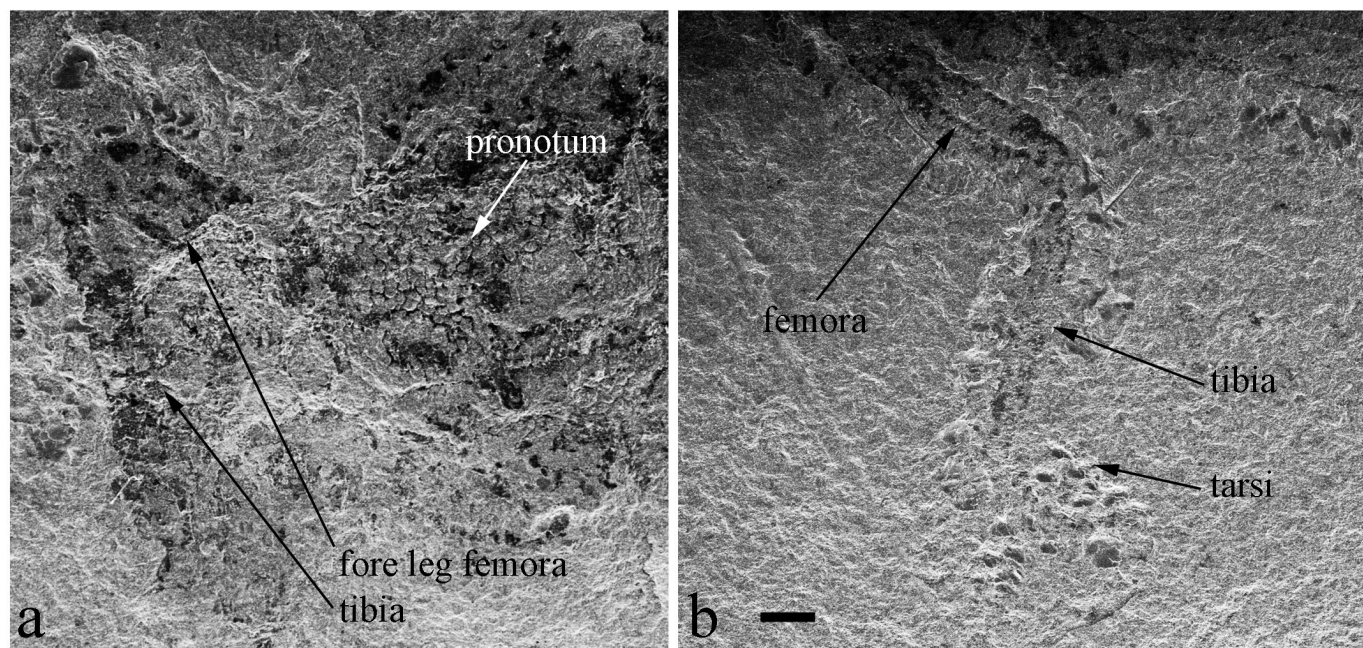




**Extended Data Figure 2 | Representative of the stem hymenopterids, *Avioxyela gallica* gen. et sp. nov., electron scanning photograph. a,** Two wings partly overlapping. **b,** Reconstruction of the two forewings. **c,** Mirror of

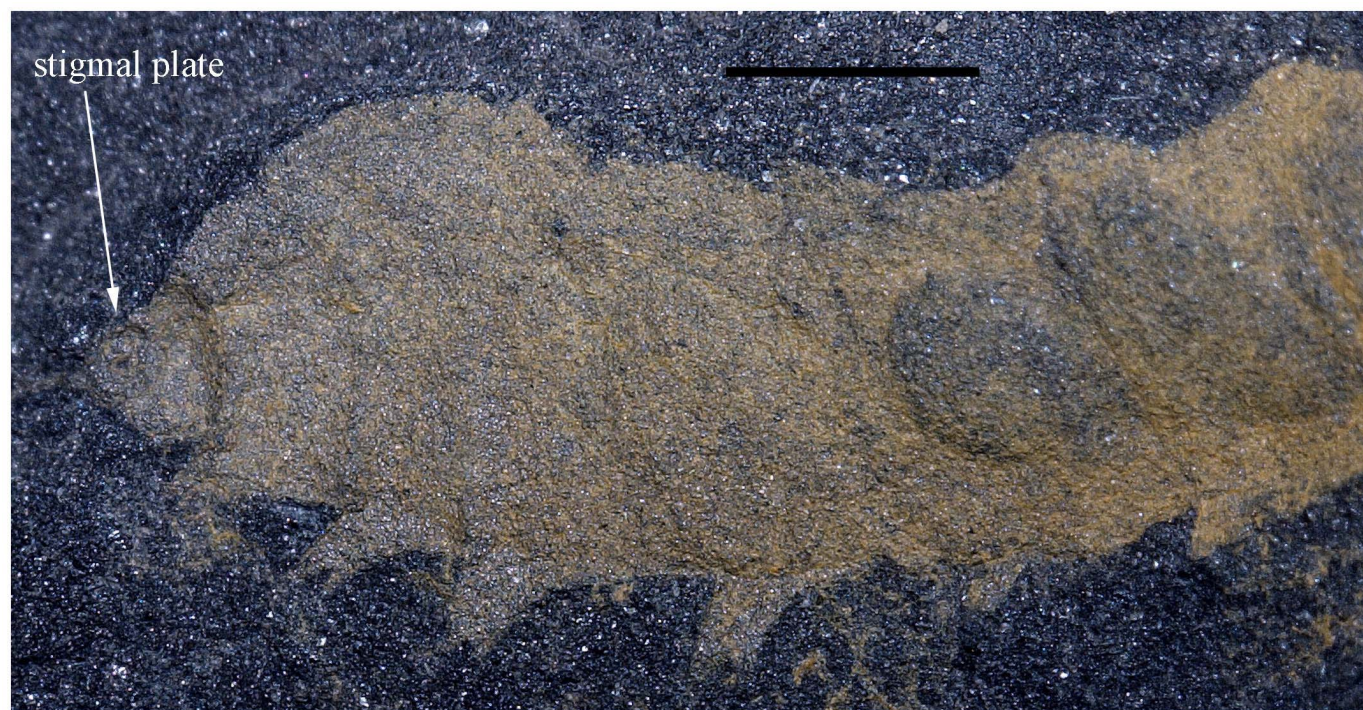
the second wing to show the identity of pattern with the first wing. **d, e,** Electron scanning photograph of radial vein of second wing with interpretation of the veins. Scale bar, 1 mm (a, b, c).





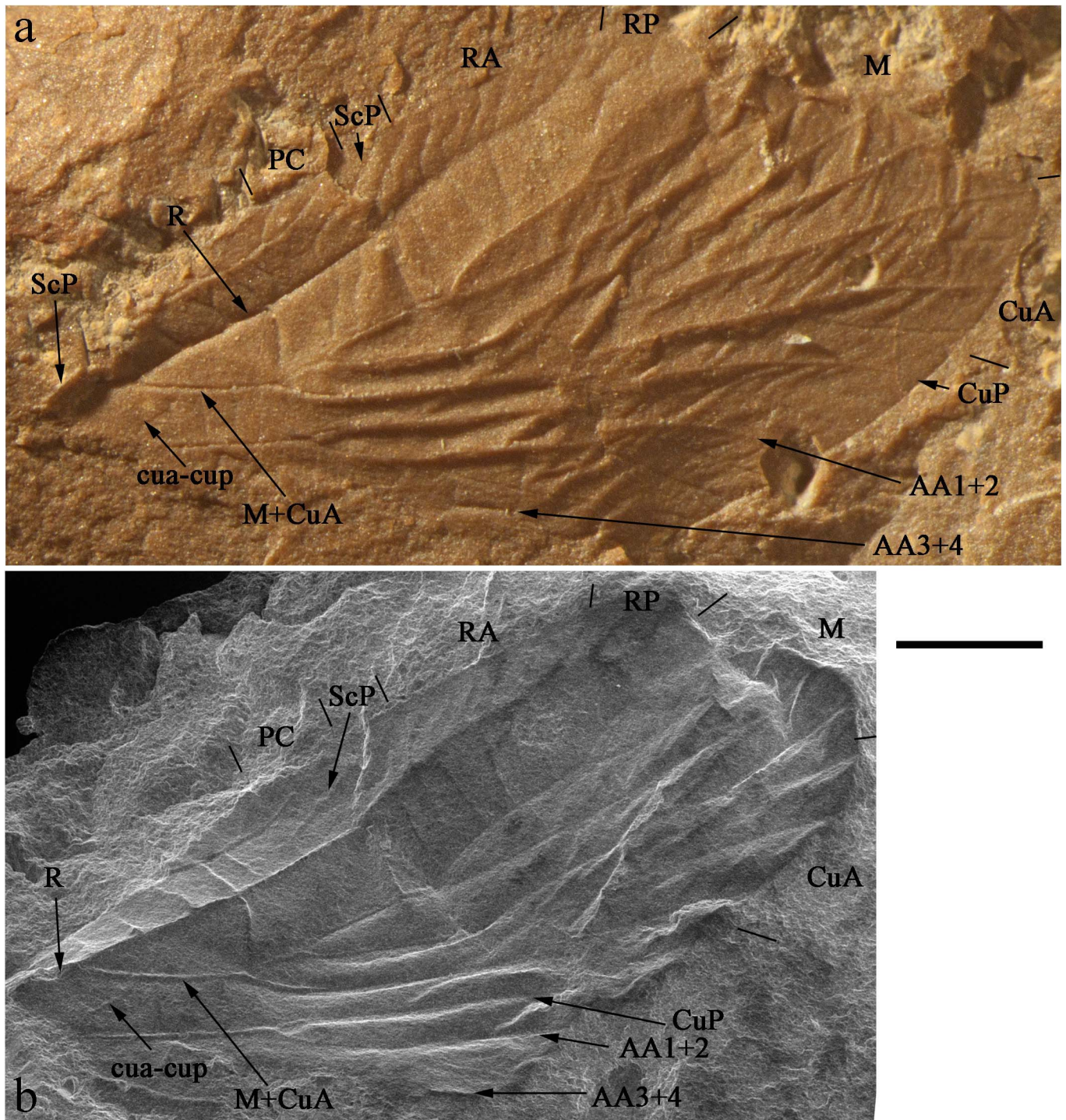
**Extended Data Figure 3** | Representative of the stem coleopterids, *Stephanastus polinae* gen. et sp. nov. Electron scanning photographs. a, Foreleg and pronotum. b, Mid-leg. Scale bar, 1 mm (a, b).





Extended Data Figure 4 | Holometabolous larva, *Metabolarva bella* gen. et sp. nov. Photograph of posterior half of body. Scale bar, 1 mm.

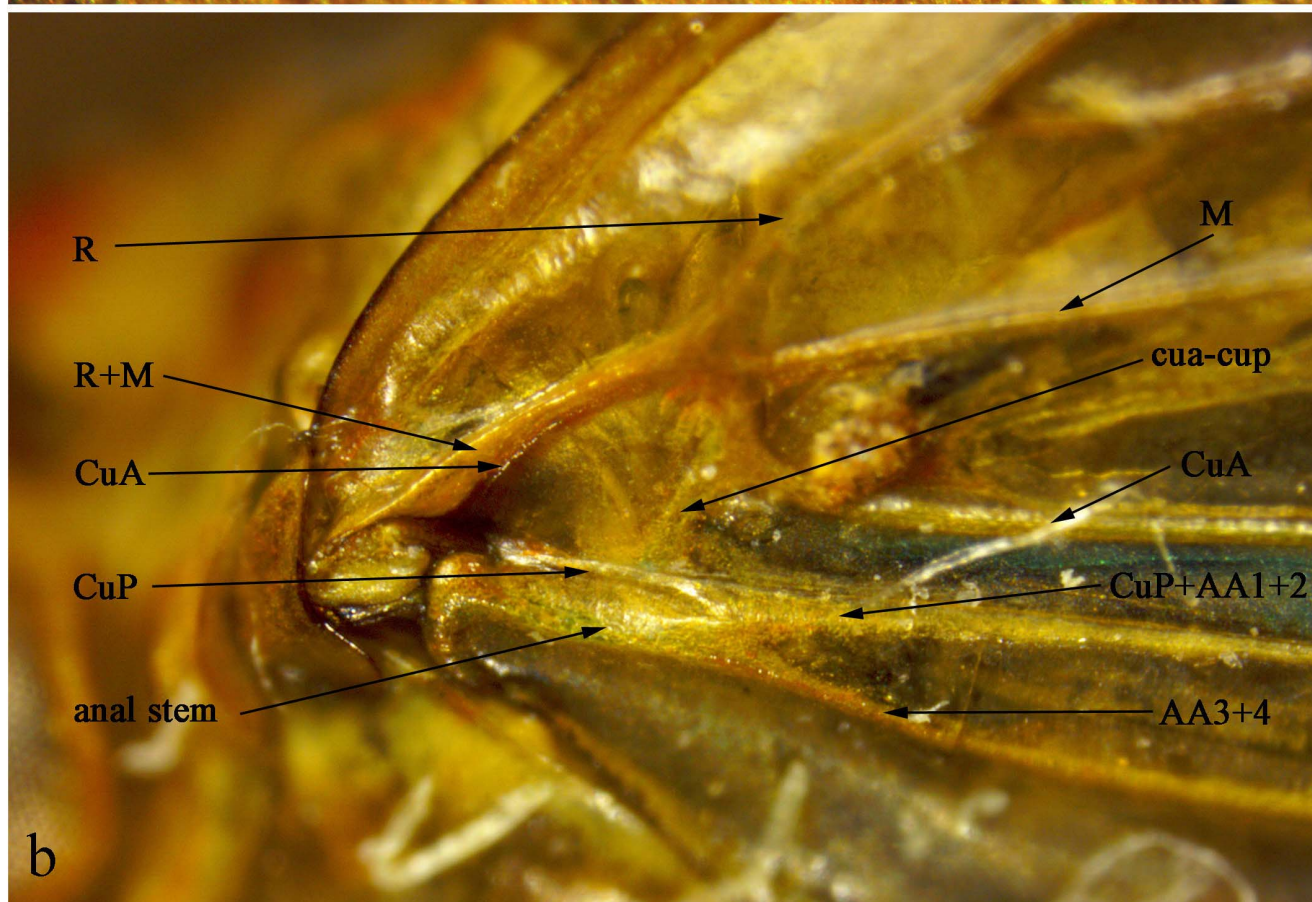
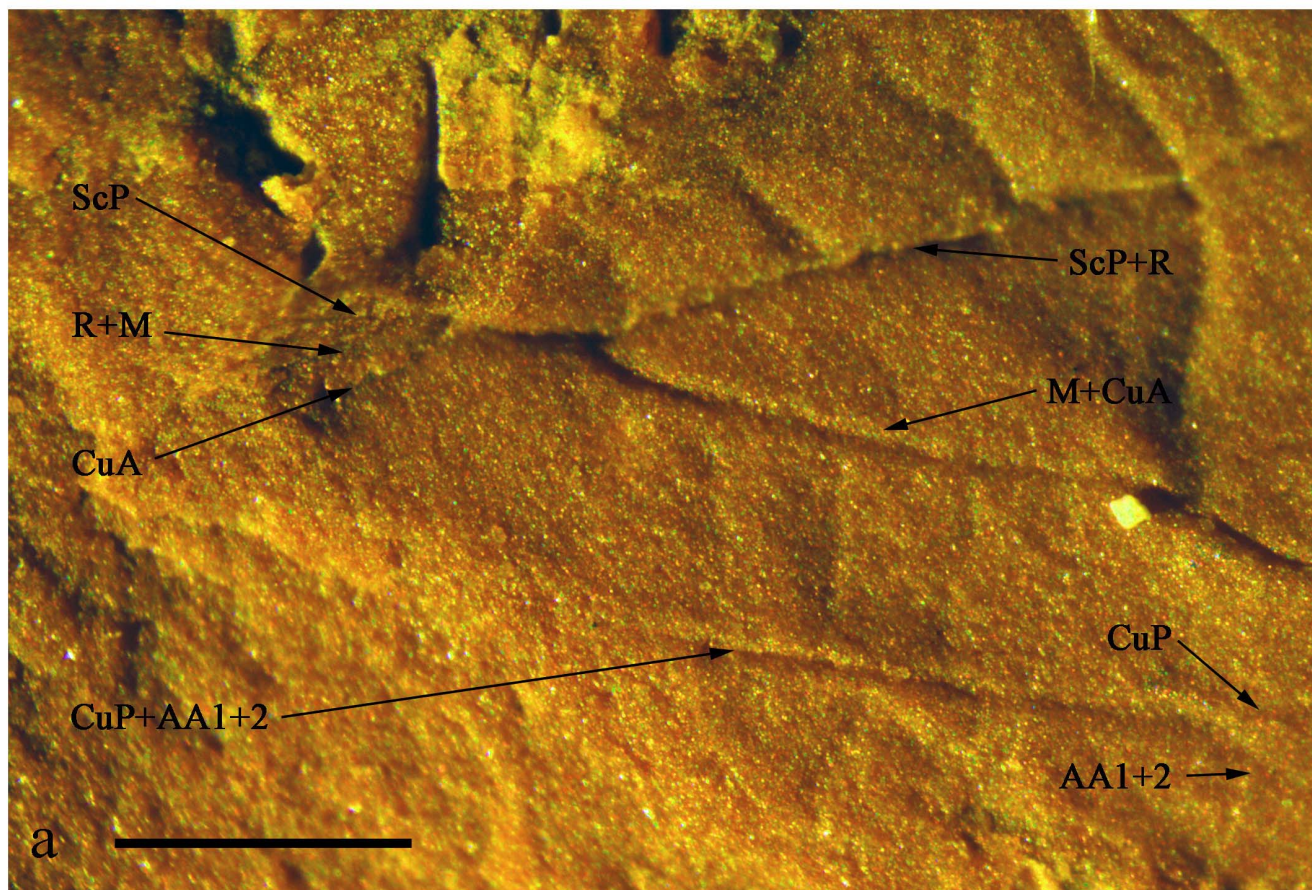




Extended Data Figure 5 | Representative of stem Euhemiptera, *Aviorrhyncha magnifica* gen. et sp. nov. a, Photograph of part. b, Electron

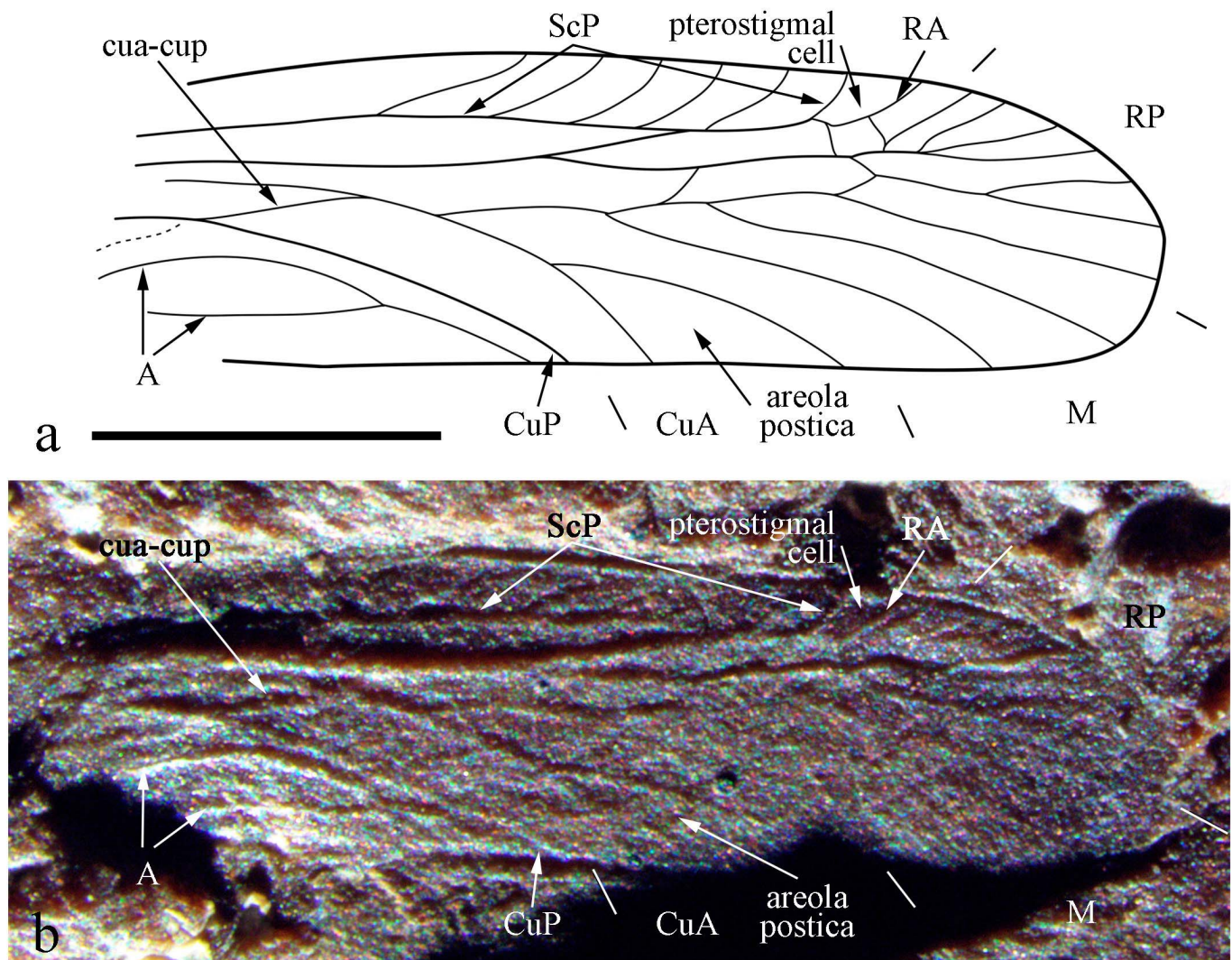
scanning photograph of counterpart. AA1 + 2, first anal anterior vein; AA3 + 4, second anal anterior vein; PC, precosta. Scale bar, 1 mm (a, b).



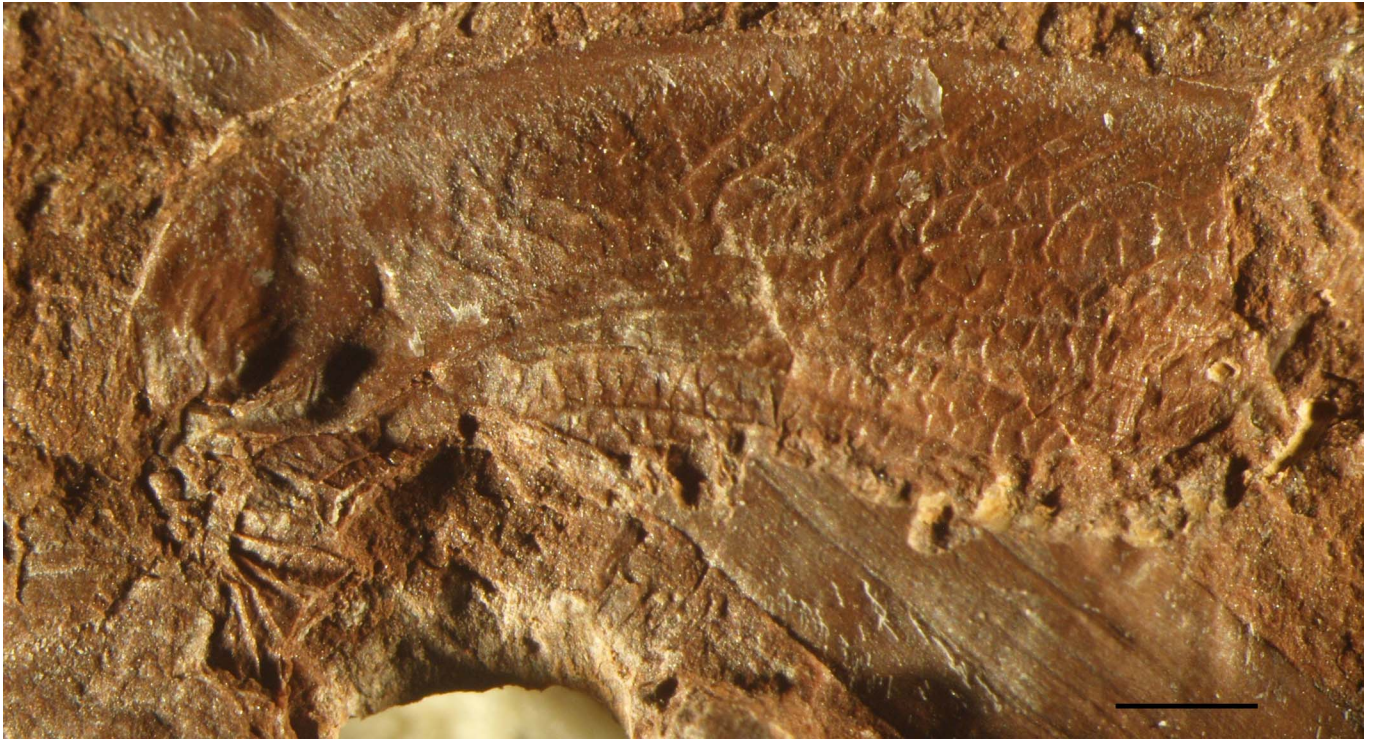




**Extended Data Figure 6 | Euhemiptera wing venation.** **a**, Wing base of *Aviorrhyncha magnifica* gen. et sp. nov. **b**, Wing base of a modern Fulgoroidea. Scale bar, 1 mm (a, b).



**Extended Data Figure 7** | Representative of stem Psocodea, *Westphalopsocus pumilio* gen. et sp. nov. **a**, Reconstruction of forewing. **b**, Photograph. A, anal veins. Scale bar, 1 mm (a, b).



Extended Data Figure 8 | Very small forewing of a dictyopteran from Avion. Specimen Avion 13. Scale bar, 1 mm.



# Feature detection and orientation tuning in the *Drosophila* central complex

Johannes D. Seelig<sup>1</sup> & Vivek Jayaraman<sup>1</sup>

Many animals, including insects, are known to use visual landmarks to orient in their environment. In *Drosophila melanogaster*, behavioural genetics studies have identified a higher brain structure called the central complex as being required for the fly's innate responses to vertical visual features<sup>1</sup> and its short- and long-term memory for visual patterns<sup>2–4</sup>. But whether and how neurons of the fly central complex represent visual features are unknown. Here we use two-photon calcium imaging in head-fixed walking and flying flies to probe visuomotor responses of ring neurons—a class of central complex neurons that have been implicated in landmark-driven spatial memory in walking flies<sup>2,3</sup> and memory for visual patterns in tethered flying flies<sup>5</sup>. We show that dendrites of ring neurons are visually responsive and arranged retinotopically. Ring neuron receptive fields comprise both excitatory and inhibitory subfields, resembling those of simple cells in the mammalian primary visual cortex. Ring neurons show strong and, in some cases, direction-selective orientation tuning, with a notable preference for vertically oriented features similar to those that evoke innate responses in flies<sup>1,2</sup>. Visual responses were diminished during flight, but, in contrast with the hypothesized role of the central complex in the control of locomotion<sup>6</sup>, not modulated during walking. Taken together, these results indicate that ring neurons represent behaviourally relevant visual features in the fly's environment, enabling downstream central complex circuits to produce appropriate motor commands<sup>6</sup>. More broadly, this study opens the door to mechanistic investigations of circuit computations underlying visually guided action selection in the *Drosophila* central complex.

Flies display a variety of visual pattern- and position-dependent behaviours, including stripe fixation<sup>2</sup>, short-term orientation memory<sup>2</sup>, pattern learning<sup>4</sup> and place learning<sup>3</sup>. Common to these behaviours is a need to detect and respond to specific features in the insect's visual surroundings. In addition, all these behaviours require the central complex<sup>1–5</sup>, a deep brain region that has also been implicated in motor control<sup>6</sup>. We used two-photon calcium imaging in genetically targeted populations of central complex input neurons in behaving flies to investigate their potential visuomotor role. We focused on the dendritic responses of ring neurons—neurons that connect the lateral triangle (LTR) to the ellipsoid body<sup>7–9</sup> (Fig. 1a) and have been specifically implicated in visuomotor memory<sup>2,3</sup>.

Electron microscopy in the locust has shown that dendrites of ring neuron analogues arborize in specialized structures in the LTR called microglomeruli, where they are contacted by axonal projections from visual areas<sup>10</sup>. Confocal images of the *Drosophila* LTR labelled with green fluorescent protein (GFP) under the control of a pan-neuronal driver line, *R57C10* (ref. 11), revealed a similar dense microglomerular substructure in the region (Fig. 1b, Supplementary Videos 1–4).

We asked if LTR microglomeruli respond to visual input. We used two-photon imaging with the calcium indicator GCaMP expressed pan-neuronally to record neural activity in the LTR of head-fixed *Drosophila* placed at the centre of a visual arena (Fig. 1c, d). Flies were presented with small bright vertical bars moving horizontally at

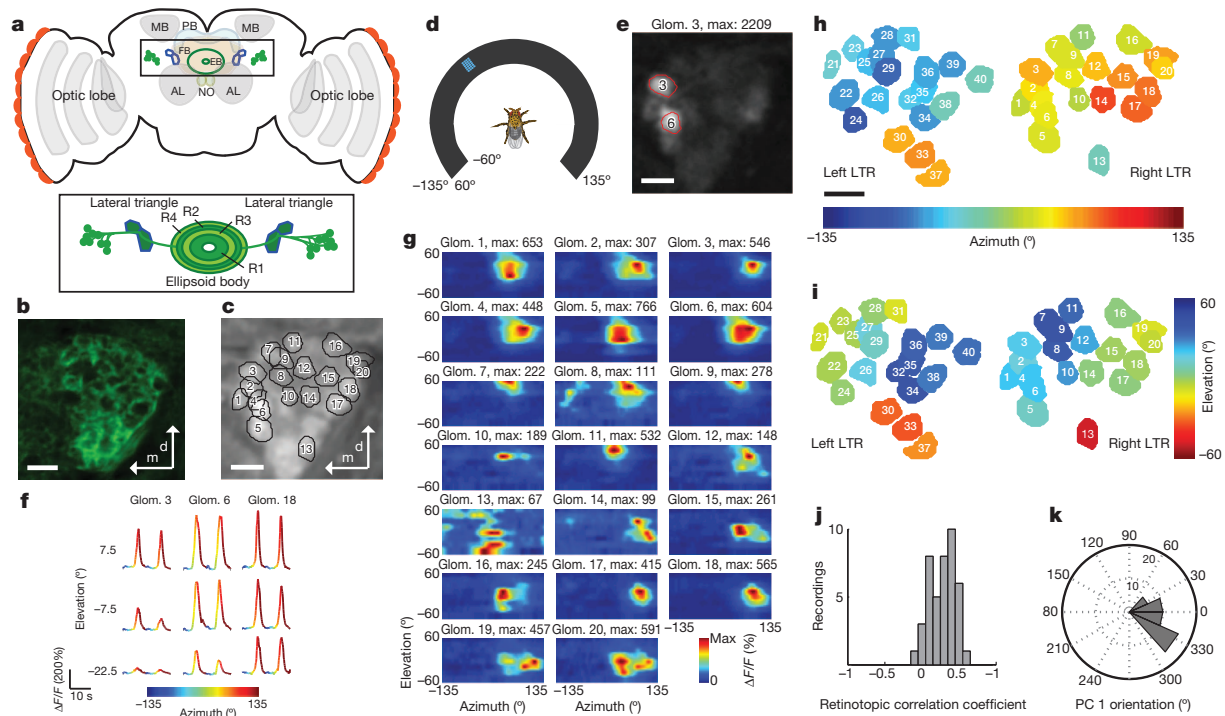
different elevations and LTR calcium transients were recorded from several planes of focus on one or both sides of the brain in a single experiment (see Methods). Calcium transients showed strong temporal correlations at the spatial scale expected of LTR microglomeruli (Fig. 1e, Supplementary Video 5 and Extended Data Fig. 1a). Visual stimuli evoked robust calcium transients in a subset of microglomeruli, but only when the localized stimuli were in specific spatial locations around the fly (Fig. 1f). We computed receptive fields for responsive microglomeruli and found that most receptive fields are centred in the ipsilateral visual hemifield (Fig. 1g–i, Extended Data Fig. 1b and Methods). Finally, LTR microglomeruli are clustered retinotopically and principal component analysis based on receptive field centres indicates that they form a spatial map with axes that are almost parallel to the fly's visual field (Fig. 1j, k and Extended Data Fig. 1c–e).

We next examined the anatomical relationship between the LTR and individual classes of ring neurons that send arborizations to the region<sup>8,12</sup>. We studied dendritic arborization patterns of ring neurons targeted by *EB1-GAL4*, which labels R2 ring neurons required for pattern memory<sup>5</sup>, and *c232-GAL4*, which labels R3/R4d neurons required for spatial memory<sup>2</sup> (Supplementary Videos 1 and 2 for R3/R4d; Supplementary Videos 3 and 4 for R2). In agreement with past anatomical work<sup>9</sup>, different ring neuron classes arborize in specific contiguous parts of the LTR (Extended Data Fig. 2a, b). Each ring neuron in these classes extends dendrites into a single microglomerulus in the LTR, and sends axonal arbors throughout a class-specific ring of the ellipsoid body (Fig. 2a, Extended Data Fig. 2c–e, see also<sup>7,12</sup>).

To understand whether different types of ring neurons have distinctive visual response properties, we mapped receptive fields for dendritic microglomeruli of R3/R4d and R2 ring neurons (Extended Data Fig. 3a–c and Extended Data Fig. 3d–f, respectively). We found visual responses in ~7/40 *c232-GAL4*-labelled microglomeruli—corresponding to ~7/20 R4d microglomeruli (Extended Data Fig. 3a)—and ~14/20 of R2 microglomeruli labelled by *EB1-GAL4*. Receptive fields for R2 and R4d neurons cover large parts of the visual field, with highest density near the midline of the ipsilateral visual field (Extended Data Fig. 3g, h). In summary, R4d and R2 microglomeruli appear to have similar visual response properties and overlapping receptive fields, but with peak sensitivity in different parts of the visual field (Extended Data Fig. 3i–k).

We next probed the fine structure of microglomerular receptive fields using sparse white noise stimuli (Fig. 2b, see Methods). Reverse correlation of microglomerular responses revealed prominent inhibitory subfields in the receptive fields (Fig. 2c for R2; Fig. 3a for R4d). The spatial scales of receptive field structure we observe is within the range for visual features that evoke strong innate responses in flies, and that are used for visual pattern learning in tethered flies<sup>2,3</sup>. To test the validity of these white-noise-based receptive fields, we used them to predict responses to novel bar stimuli (Fig. 2d, see Methods). The predicted responses captured much of the temporal and spatial variation in the data (Fig. 2e), with high correlations between estimated and actual responses (Extended Data Fig. 4a, b).

<sup>1</sup>Janelia Farm Research Campus, Howard Hughes Medical Institute, 19700 Helix Drive, Ashburn, Virginia 20147, USA.



**Figure 1 | *Drosophila* ellipsoid body ring neurons arborize in visually responsive LTR microglomeruli which show a retinotopic organization.** **a**, Schematic of fly central brain showing antennal lobe (AL), mushroom bodies calyces (MB) and optic lobes along with sub-structures of the central complex: ellipsoid body (EB), fan-shaped body (FB), protocerebral bridge (PB) and noduli (NO). Inset shows EB ring neurons (R1–R4) and LTR. **b**, Frame of a confocal stack showing LTR microglomeruli labelled by pan-neuronal GFP expression. **c**, Projection of two-photon calcium imaging video of LTR with overlay of microglomeruli selected based on responses to moving visual stimuli. In individual planes of focus,  $20 \pm 5$  microglomeruli were delineated; over multiple planes of focus,  $30 \pm 4$  microglomeruli were delineated ( $n = 11$  flies) (see Methods). **d**, Schematic of receptive field mapping set-up with fly positioned in centre of curved visual display. **e**, Sample frame from trial

showing responses in selected glomeruli (red outlines). **f**, Calcium transients of three LTR microglomeruli in response to visual stimulus moving left to right in front of the fly at different elevations. **g**, Two-dimensional response maps (two-trial averages) for all microglomeruli shown in **c**. **h**, **i**, LTR microglomeruli from left and right hemisphere of same fly, coloured according to the centre of the receptive field in azimuth (**h**) and elevation (**i**). **j**, Histogram of correlation coefficients between receptive field centre and anatomical position. For  $n = 42$  focal planes with  $20 \pm 5$  glomeruli, correlation is significantly different than for randomly arranged microglomeruli ( $r = 0.27 \pm 0.17$ ,  $P = 2.3 \times 10^{-17}$ ,  $n = 11$  flies), indicating retinotopy in the organization of microglomeruli across flies. **k**, Histogram of primary retinotopic axis of LTR map as found by principal component analysis (see Methods,  $n = 42$  focal planes, 11 flies). All scale bars, 5  $\mu\text{m}$ .

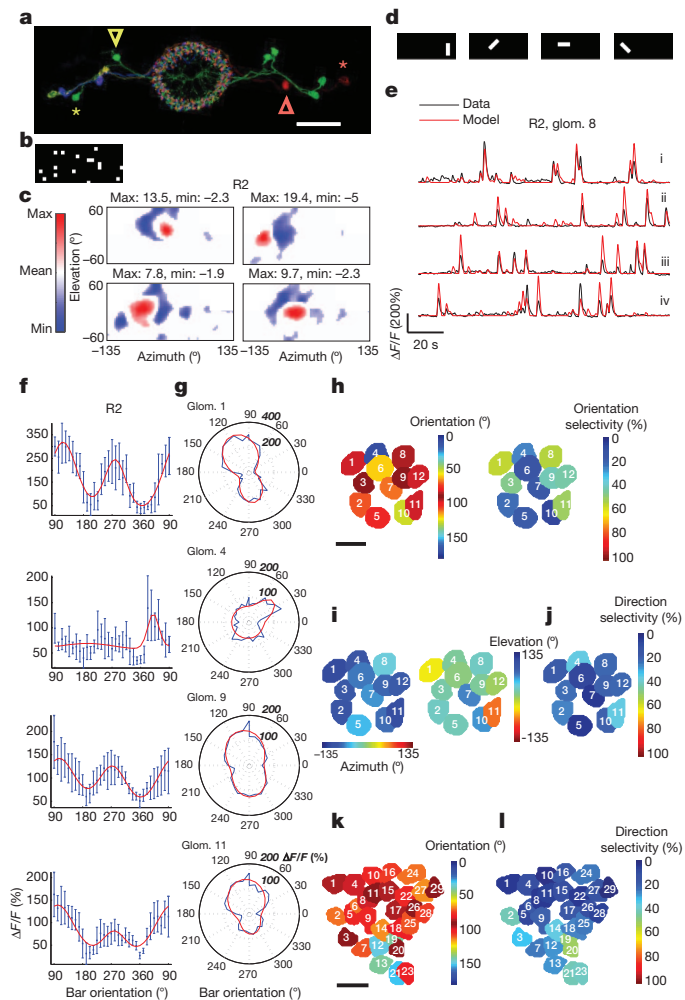
Noting that the receptive field structure of ring neuron inputs resembles those of simple cells in mammalian primary visual cortex<sup>13</sup>, we next asked if these neurons share other response properties. Indeed, when we presented flies with a series of moving oriented bars, we found strong orientation tuning in microglomerular response patterns (Fig. 2f, g). As expected for receptive field structures with both excitatory and inhibitory lobes, microglomeruli also showed orientation tuning when presented with bars of opposite contrast, that is, dark bars on a bright background (Extended Data Fig. 5a, b), as are often used in fly behavioural studies<sup>1,2</sup>. Examining orientation tuning across the population of microglomeruli, we observed a strong preference for vertically oriented bars (Fig. 2h and Extended Data Fig. 4h, i for R2; Extended Data Fig. 4c, k, l for R4d; Fig. 2k and Extended Data Fig. 4g, n, o for the pan-neuronal line). Receptive fields tuned to vertical orientations are distributed across the visual field (Fig. 2i, j for R2; Extended Data Fig. 4c, d for R4d; Extended Data Fig. 4f for the pan-neuronal line). A fraction of neurons also shows direction-selectivity (Fig. 2j and Extended Data Fig. 4j for R2; Extended Data Fig. 4e, m for R4d; Fig. 2l and Extended Data Fig. 4p for the pan-neuronal line).

We next assessed the degree of stereotypy in response properties of ring neuron dendrites of different flies. We found strong correlations across flies in receptive field structure for R4d (Fig. 3a, b, Extended Data Figs 6 and 8), and R2 (Fig. 3c, Extended Data Figs 7 and 8).

Ring neurons and the ellipsoid body have often been ascribed a role in complex visuomotor tasks<sup>2,3</sup>. We examined the possible motor function of ring neurons by assessing potential correlations between neural activity and locomotion in tethered flies walking on an air-supported

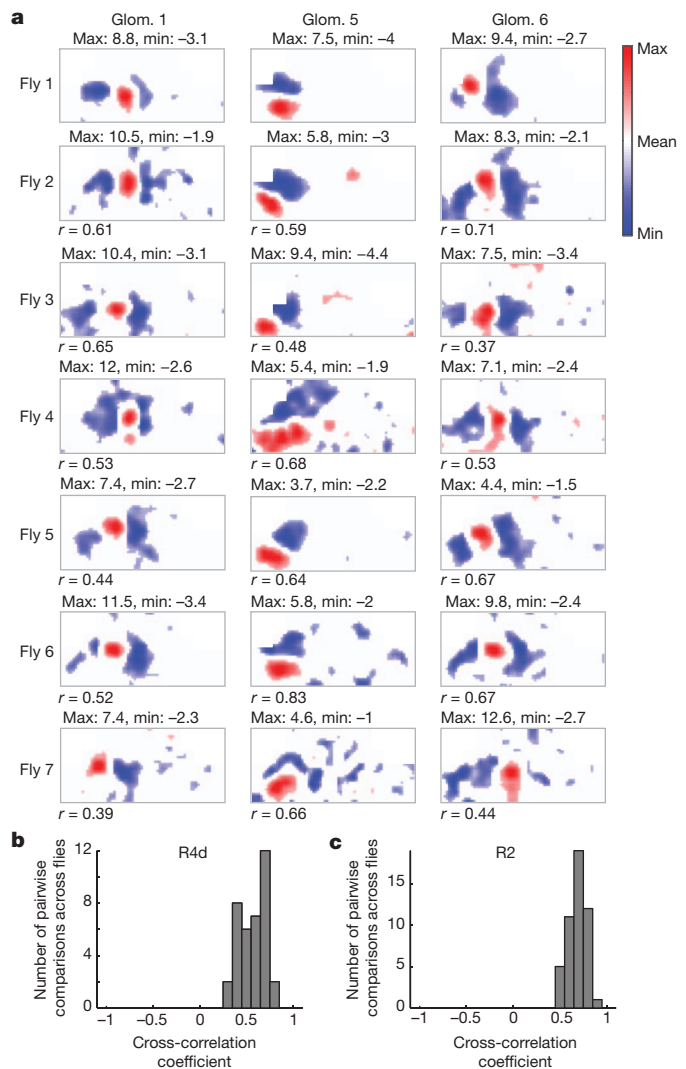
ball<sup>14</sup> or flying<sup>15</sup>, in darkness or in the presence of a bright stripe moving left or right in front of the fly (Fig. 4a, see also Methods and Supplementary Videos 6 and 7). Although some R3/R4d neurons did show occasional correlations with locomotion when flies walked in the dark, and responses from visually stimulated animals showed occasional modulation during walking, the changes were within the expected variability of visual responses in stationary flies (Fig. 4b–d). Responses were also insensitive to walking direction. Overall, LTR visual responses could be modelled accurately without taking walking state into account (Extended Data Fig. 9a–e and Extended Data Fig. 10). By contrast, responses to visual stimuli were consistently diminished during tethered flight (Fig. 4e, f, Extended Data Fig. 9f–h and Supplementary Video 7), but showed no obvious correlations with flight direction as assessed by differences in wingbeat amplitude envelope. Thus, ring neuron LTR responses were modulated by motor state, but not in a manner consistent with a direct role in motor coordination, and in a markedly different manner than in the optic lobe<sup>15,16</sup>.

Behavioural genetics studies in *Drosophila* have suggested that the central complex is required for a wide range of important sensorimotor functions. However, in the absence of physiological recordings from the region in flies, it has been challenging to determine its role in the diversity of behaviours that it has been implicated in. We studied the visuomotor responses of ring neurons, which provide input from the LTR to the ellipsoid body of the central complex. Analogous neurons in other insects respond to polarized and unpolarized light<sup>17–19</sup>, and to mechanosensory stimulation<sup>20</sup>, one of the sensory modalities that we did not explore but which may partly account for unresponsive



**Figure 2 | Ring neurons are tuned to specific visual features and orientations.** **a**, Multicoloured FLP-out of R2 neurons showing three cell bodies on each side along with their colour-matched microglomeruli (green, light green and purple at left; two green and one red at right). Red and yellow asterisks mark two cell bodies (one on each side, lateral) and arrows of like colour their respective LTR microglomeruli (medial). All neurons send processes throughout ellipsoid body rings. Scale bar, 30  $\mu$ m. **b**, Example frame of white noise stimulus used for receptive field mapping using reverse correlation. **c**, Sample receptive fields of R2 microglomeruli. Red subfields show excitatory responses  $>30\%$  of maximum; blue subfields show inhibitory responses  $<30\%$  of minimum of mean-subtracted weighted average. See Extended Data Fig. 8 for all receptive fields. **d**, Bright bars with four different orientations used as test features. **e**, Modelled and actual (black)  $\Delta F/F$  changes of an R2 microglomerulus in response to differently oriented bars (fly 2 in Extended Data Fig. 7). In red is shown the trial used for fitting parameters (i), and the tests (ii–iv). **f**, Orientation tuning curves for R2 neurons (two-trial average, fit in red). The measurement  $90^\circ$  corresponds to back-to-front movement of the vertical bar,  $270^\circ$  to front-to-back movement of the vertical bar. Error bars represent standard deviation. **g**, Polar plots of orientation tuning data and fits (red) for data shown in **f**. Numbers italicized in bold indicate  $\Delta F/F$  (%). **h**, Microglomeruli of R2 neurons coloured by orientation preference (collapsed to  $0^\circ$ – $180^\circ$ ) and orientation selectivity (two-trial average). **i**, Same microglomeruli as in **h** coloured by azimuth and elevation of centre of their excitatory receptive fields, measured using horizontally moving bars as described in Extended Data Fig. 1 (two-trial average). **j**, Direction selectivity of same microglomeruli (two-trial average). **k**, **l**, Preferred orientation (collapsed to  $0^\circ$ – $180^\circ$ ) (**k**) and direction selectivity (**l**) of microglomeruli in pan-neuronal line (both four-trial average). See Methods for analysis details. Scale bars for **h** and **k**, 5  $\mu$ m.

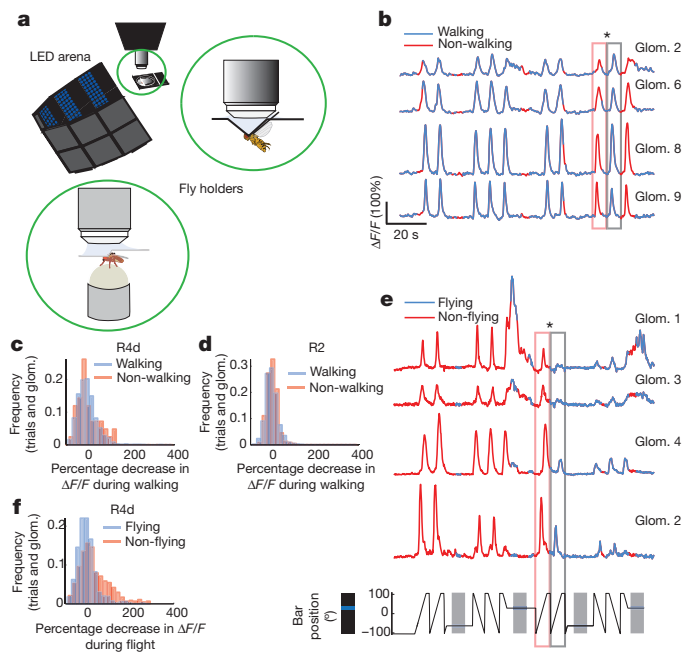
neurons in our study. We found that R2 and R4d ring neurons are visually responsive, and these responses are not significantly modulated by walking state. Although visual responses are diminished during



**Figure 3 | Ring neuron LTR microglomeruli show stereotyped receptive field properties across flies.** **a**, Subset of receptive fields measured in R4d neurons across seven flies aligned by similarity (see Extended Data Fig. 8b for full set). Numbers below receptive fields are cross-correlations with top receptive field in the column as template. **b**, Histogram of cross-correlation values calculated for R4d neuron receptive fields with best-matched template. **c**, Histogram of cross-correlation values for R2 neurons ( $n = 6$  flies, see Extended Data Fig. 8a for receptive fields).

flight, they do not vary systematically with wingbeat patterns associated with turns. It is possible that outputs of these ring neurons within the ellipsoid body rings could be more sensitive to motor actions, but our physiological results in their dendrites during behaviour are inconsistent with a major role for these ring neurons in motor coordination in the fly. Further, the high degree of stereotypy that we observe in ring neuron receptive field characteristics across flies indicates that, rather than directly performing motor coordination, these neurons probably provide downstream central complex circuits with similar behaviourally relevant visual feature sets on which to base motor decisions. As a striking example, the strong vertical tuning preference we observe in the LTR may partly underlie the tendency of flies to fixate on vertical edges during both flight and walking<sup>1,2,21</sup>. Recent work has suggested that vertical stripe fixation during walking largely relies on a hypothesized position system sensitive to local luminance changes that can operate independently of neural circuits involved in optomotor responses to wide-field motion<sup>21,22</sup>. The response properties of R2 and R4d neurons is consistent with a role in such a position system<sup>22</sup>.





**Figure 4 | Ring neuron visual responses are not significantly modulated during walking but diminished during flight.** **a**, Set-up for two-photon imaging in behaving flies. Insets show schematic of fly tethered in flying holder and positioned on air-supported ball in walking holder. **b**, Subset of simultaneously recorded R2 neurons during walking. Boxes indicated with an asterisk show responses to identical visual stimuli when fly is stationary versus walking. Azimuthal position of visual stimulus shown in **e** (bottom). **c**, Distributions of R4d neuron visual responses during walking and non-walking conditions are not significantly different ( $n = 14$  flies; trials<sub>walking</sub> = 1,722; trials<sub>walking <50%</sub> = 42; mean<sub>walking</sub> =  $0 \pm 44.3\%$ ; mean<sub>walking <50%</sub> =  $1.1 \pm 49.7\%$ ;  $P = 0.45$ ). Percentage decrease in  $\Delta F/F$  during walking was calculated as  $(\Delta F/F - (\Delta F/F)_{\text{walk}}) / (\Delta F/F)_{\text{walk}}$ . **d**, Same as **c** for R2 neurons ( $n = 8$  flies; trials<sub>walking</sub> = 2,015; trials<sub>walking <50%</sub> = 245; mean<sub>walking</sub> =  $0 \pm 28.1\%$ ; mean<sub>walking <50%</sub> =  $-2.2 \pm 24.4\%$ ;  $P = 0.37$ ). **e**, Subset of simultaneously recorded R4d microglomeruli during flight. Boxes indicated with an asterisk show diminished responses to identical visual stimuli during flight. **f**, Distributions of all R4d microglomeruli recorded shows significant shift towards lower responses during flight ( $n = 13$  flies; trials<sub>flying</sub> = 759; trials<sub>flying <50%</sub> = 481; mean<sub>flying</sub> =  $0 \pm 42.2\%$ ; mean<sub>flying <50%</sub> =  $31.2 \pm 72.3\%$ ;  $P = 6 \times 10^{-15}$ ). Percentage decrease in  $\Delta F/F$  during flying was calculated as  $(\Delta F/F - (\Delta F/F)_{\text{fly}}) / (\Delta F/F)_{\text{fly}}$ . All  $P$ -values show two-sample Kolmogorov–Smirnov test.

The retinotopic bias, structure of excitatory and inhibitory subfields, orientation tuning and direction selectivity we see are reminiscent of those seen in calcium imaging studies in simple cells in mammalian visual cortex<sup>13</sup>, providing an interesting example of how evolutionarily distant visual systems with different types of eyes nonetheless use similar feature sets to process visual scenes. From Hubel and Wiesel's findings several decades ago<sup>23</sup> to the present, significant progress has been made on identifying neural representations used at different stages in the mammalian visual system. However, mechanisms underlying simple cell responses are not yet fully understood<sup>24</sup>. With its array of genetic tools, *Drosophila melanogaster* may allow us to uncover how spatiotemporal interactions of excitatory and inhibitory inputs might produce similar orientation tuning and direction selectivity in the LTR<sup>25,26</sup>. There is considerable evidence for spatially tuned visual responses in the lobula complex and anterior optic tubercle in other insects<sup>27–30</sup>, suggesting that components of ring neuron input response properties may also arise from selective averaging of weaker and more broadly distributed tuning preferences in such areas.

Overall, our findings lay the groundwork for future research into how this genetic model organism's small brain uses feature and pattern information for visual orientation and navigation.

## METHODS SUMMARY

Calcium imaging experiments to measure receptive fields were performed with *UAS-GCaMP3;c232-GAL4*, *UAS-GCaMP3;EB1-GAL4*, and *pJFRC7-20XUAS-GCaMP5.003 (VK00005)/R57C10-GAL4* flies. We used *pGP-JFRC7-20XUAS-IVS-GCaMP6f 15.693(attP40)/R57C10-GAL4* flies for pan-neuronal orientation tuning experiments and *pGP-JFRC7-20XUAS-IVS-GCaMP6s 15.641 (attP40)/c232-GAL4* for orientation tuning experiments with dark bars. *c232-GAL4* and *EB1-GAL4* were gifts from M. Heisenberg and T. Lee, respectively.

For calcium imaging experiments, female flies were mounted in custom stainless steel shim holders with designs similar to those previously described<sup>14,15</sup>. The dissection procedure, LED visual arena and two-photon imaging set-up were similar to those used in previous experiments<sup>14</sup>. All data analysis was performed in MATLAB (MathWorks).

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 March; accepted 22 August 2013.

Published online 9 October 2013.

- Bausenwein, B., Muller, N. R. & Heisenberg, M. Behavior-dependent activity labeling in the central complex of *Drosophila* during controlled visual stimulation. *J. Comp. Neurol.* **340**, 255–268 (1994).
- Neuser, K., Triphan, T., Mronz, M., Poeck, B. & Strauss, R. Analysis of a spatial orientation memory in *Drosophila*. *Nature* **453**, 1244–1247 (2008).
- Ofstad, T. A., Zuker, C. S. & Reiser, M. B. Visual place learning in *Drosophila melanogaster*. *Nature* **474**, 204–207 (2011).
- Liu, G. *et al.* Distinct memory traces for two visual features in the *Drosophila* brain. *Nature* **439**, 551–556 (2006).
- Pan, Y. *et al.* Differential roles of the fan-shaped body and the ellipsoid body in *Drosophila* visual pattern memory. *Learn. Mem.* **16**, 289–295 (2009).
- Strauss, R. The central complex and the genetic dissection of locomotor behaviour. *Curr. Opin. Neurobiol.* **12**, 633–638 (2002).
- Young, J. M. & Armstrong, J. D. Structure of the adult central complex in *Drosophila*: organization of distinct neuronal subsets. *J. Comp. Neurol.* **518**, 1500–1524 (2010).
- Hanesch, U., Fischbach, K. F. & Heisenberg, M. Neuronal architecture of the central complex in *Drosophila melanogaster*. *Cell Tissue Res.* **257**, 343–366 (1989).
- Renn, S. C. P. *et al.* Genetic analysis of the *Drosophila* ellipsoid body neuropil: organization and development of the central complex. *J. Neurobiol.* **41**, 189–207 (1999).
- Träger, U., Wagner, R., Bausenwein, B. & Homberg, U. A novel type of microglomerular synaptic complex in the polarization vision pathway of the locust brain. *J. Comp. Neurol.* **506**, 288–300 (2008).
- Jenett, A. *et al.* A GAL4-driver line resource for *Drosophila* neurobiology. *Cell Rep.* **2**, 991–1001 (2012).
- Wang, J., Zugates, C. T., Liang, I. H., Lee, C. H. J. & Lee, T. M. *Drosophila* Dscam is required for divergent segregation of sister branches and suppresses ectopic bifurcation of axons. *Neuron* **33**, 559–571 (2002).
- Bonin, V., Histed, M. H., Yurgenson, S. & Reid, R. C. Local diversity and fine-scale organization of receptive fields in mouse visual cortex. *J. Neurosci.* **31**, 18506–18521 (2011).
- Seelig, J. D. *et al.* Two-photon calcium imaging from head-fixed *Drosophila* during optomotor walking behavior. *Nature Methods* **7**, 535–540 (2010).
- Maimon, G., Straw, A. D. & Dickinson, M. H. Active flight increases the gain of visual motion processing in *Drosophila*. *Nature Neurosci.* **13**, 393–399 (2010).
- Chiappe, M. E., Seelig, J. D., Reiser, M. B. & Jayaraman, V. Walking modulates speed sensitivity in *Drosophila* motion vision. *Curr. Biol.* **20**, 1470–1475 (2010).
- Heinze, S. & Reppert, S. M. Sun compass integration of skylight cues in migratory monarch butterflies. *Neuron* **69**, 345–358 (2011).
- Heinze, S. & Homberg, U. Maplike representation of celestial E-vector orientations in the brain of an insect. *Science* **315**, 995–997 (2007).
- Rosner, R. & Homberg, U. Widespread sensitivity to looming stimuli and small moving objects in the central complex of an insect brain. *J. Neurosci.* **33**, 8122–8133 (2013).
- Guo, P. & Ritzmann, R. E. Neural activity in the central complex of the cockroach brain is linked to turning behaviors. *J. Exp. Biol.* **216**, 992–1002 (2013).
- Bahl, A., Ammer, G., Schilling, T. & Borst, A. Object tracking in motion-blind flies. *Nature Neurosci.* **16**, 730–738 (2013).
- Poggio, T. & Reichard, W. Theory of pattern induced flight orientation of fly *Musca domestica*. *Kybernetik* **12**, 185–203 (1973).
- Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol. (Lond.)* **160**, 106–154 (1962).
- Priebe, N. J. & Ferster, D. Mechanisms of neuronal computation in mammalian visual cortex. *Neuron* **75**, 194–208 (2012).
- Freifeld, L., Clark, D. A., Schnitzer, M. J., Horowitz, M. A. & Clandinin, T. R. GABAergic lateral interactions tune the early stages of visual processing in *Drosophila*. *Neuron* **78**, 1075–1089 (2013).
- Egelhaaf, M. & Borst, A. A look into the cockpit of the fly: visual orientation, algorithms, and identified neurons. *J. Neurosci.* **13**, 4563–4574 (1993).

27. Pfeiffer, K., Kinoshita, M. & Homberg, U. Polarization-sensitive and light-sensitive neurons in two parallel pathways passing through the anterior optic tubercle in the locust brain. *J. Neurophysiol.* **94**, 3903–3915 (2005).
28. Mu, L., Ito, K., Bacon, J. P. & Strausfeld, N. J. Optic glomeruli and their inputs in *Drosophila* share an organizational ground pattern with the antennal lobes. *J. Neurosci.* **32**, 6061–6071 (2012).
29. Collett, T. Visual neurones for tracking moving targets. *Nature* **232**, 127–130 (1971).
30. O'Carroll, D. Feature-detecting neurons in dragonflies. *Nature* **362**, 541–543 (1993).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank A. Nern for sharing flies and technical advice; B. Pfeiffer and G. Rubin for providing pJFRC64, pJFRC118, and pJFRC119 flies; and R. Harris for building fly lines, sharing preliminary results, and providing critical feedback on the use

of fly reagents. We thank K. Hibbard, D. Hall, J. Kao and the Janelia Fly Core for fly crosses and support; E. Chiappe for initial anatomy experiments; M. Reiser and J. Truman for sharing equipment; J. Liu for technical support; V. Iyer for ScanImage support; T. Adelman for helpful suggestions; and L. Looger and the Janelia GENIE team for GCaMP5 and GCaMP6. We are grateful to A. Leonardo, M. Reiser, E. Chiappe, J. Freeman, A. Karpova, S. Huston and members of Vivek's laboratory for useful discussions and comments on the manuscript. This work was supported by the Howard Hughes Medical Institute.

**Author Contributions** Both authors designed the study and wrote the manuscript. J.S. carried out the experiments and data analysis.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to V.J. ([vivek@janelia.hhmi.org](mailto:vivek@janelia.hhmi.org)).

## METHODS

**Fly stocks.** All experiments were performed with female flies, with ages chosen based on expression levels of relevant fluorescent proteins. At least six animals were used for any single condition tested—specific sample sizes are noted for each set of experiments and were chosen based on the level of variability observed in initial experiments. Flies were randomly picked from their housing vials for all experiments. We used all data that passed a quality threshold based on the observed health of the fly during an experiment and the signal-to-noise ratio of the imaging signal.

Calcium imaging experiments to measure receptive fields were performed with *UAS-GCaMP3;c232-GAL4*, *UAS-GCaMP3;EB1-GAL4*, and *pJFRC7-20XUAS-GCaMP5.003 (VK00005)/R57C10-GAL4* flies. We used *pGP-JFRC7-20XUAS-IVS-GCaMP6f 15.693(attP40)/R57C10-GAL4* flies for pan-neuronal orientation tuning experiments (2–5-days old) and *pGP-JFRC7-20XUAS-IVS-GCaMP6s 15.641 (attP40)/c232-GAL4* for orientation tuning experiments (1–2-weeks old) with dark bars. *c232-GAL4* and *EB1-GAL4* were gifts from M. Heisenberg and T. Lee, respectively.

To label dendrites and axons in different ring neurons, *EB1-GAL4* and *c232-GAL4* were each crossed to *pJFRC67-3XUAS-IVS-Syt-GFP(attP18)*<sup>31,32</sup>, *pJFRC118-10XUAS-DenMark(attP40)*<sup>32,33</sup> and *pJFRC119-10XUAS-IVS-myr::TopHat2(VK00005)* (gift of B. Pfeiffer and R. Harris, unpublished).

To compare the expression patterns of *EB1-GAL4* and *c232-GAL4* to the pan-neuronal line *R57C10* in the same fly, *EB1-GAL4* and *c232-GAL4* were each crossed to *pJFRC19-13XLexAop2-IVS-myr::GFP (su(Hw)attP8)*<sup>32</sup>, *pJFRC21-10XUAS-IVS-mCD8::RFP (attP18)*<sup>32</sup>, *Sco/Cyo*; *R57C10-LexA::p65 (attP2)*<sup>34</sup>.

For stochastic single cell labelling of *EB1-GAL4* and *c232-GAL4* with three colours a 'flip-out'-based approach<sup>35</sup> (Nern *et al.*, manuscript in preparation) was used. In brief, heat-shock induced expression of FLP recombinase was used to excise FRT-flanked interruption cassettes from UAS reporter constructs carrying HA, V5 and Flag epitope tags, and stained with epitope-tag specific antibodies. This labelled a subset of the cells in the expression pattern with a stochastic combination of the three labels.

**Anatomy: fly dissections, immunohistochemistry and confocal imaging.** Confocal stacks were recorded with a  $\times 40$  or  $\times 63$  objective on a Zeiss confocal microscope. Dissections and staining were performed as previously described<sup>36,37</sup>. The primary antibody mixture consisted of 1:1,000 sheep anti-GFP (AbD Serotec), 1:1,000 rabbit anti-DsRed (Clontech), 1:100 rat anti-HA (for Syt/DenMark/HA staining, Roche), and 10% normal donkey serum (Jackson ImmunoResearch) in PBS-TX.

**Fly preparation for receptive field mapping.** Receptive field measurements were performed using a preparation described previously<sup>14</sup>, but with the behavioural apparatus removed to maximize the fly's visual field. The fly was briefly anaesthetized on ice and transferred to a cold plate held at 4 °C. The proboscis of the fly was fixed by either pressing it into the head and fixing it with wax, or by stretching it with a pair of tweezers mounted on a micromanipulator and then fixing it with a mixture of wax and colophony. We additionally removed either the front legs or all legs for receptive field measurements. The fly was glued to a pin and positioned in the holder using a micromanipulator and fixed in the holder with UV gel<sup>14</sup>. An opening was cut into the head to obtain optical access to the brain. To stop brain movement due to pulsation of muscle M16, we cut the muscle or the nerves innervating the muscles with dissection needles. The fly holder (including the micromanipulator) was then transferred to the microscope and mounted using magnetic mounts. Flies were dark adapted for 5 to 10 min before recordings started.

**Fly preparation for imaging during walking.** For walking experiments, an air-supported ball was positioned under the fly with a three-axis micromanipulator as described previously<sup>14</sup> and the walking velocity of the fly was monitored using a camera system.

**Fly preparation for imaging during tethered flight.** For flying fly experiments, a holder similar to those described previously<sup>14,15,38</sup> was used. The holder was made out of two pieces of stainless steel shim (thickness 12.7  $\mu\text{m}$ ). The shim is cut using a laser mill and then folded into its pyramidal shape. One piece of the holder is glued onto a liquid chamber similar to the one used for walking behaviour using epoxy<sup>14</sup>. After removing the front legs of the fly to prevent it from grabbing the holder, the fly is glued to a pin and inserted into the holder with a micromanipulator. The second piece of the holder is then inserted to close the pyramidal shape around the fly's head.

The setup for combined imaging and flight behaviour was similar to one previously described<sup>15</sup>. The fly was illuminated with infrared light from below. The wing beat of the fly was recorded using a mirror placed beneath the fly and a camera (Basler 602f, operating at 100 or 150 Hz frame rate)<sup>15</sup>.

**Two-photon imaging.** Calcium imaging was performed using a custom-built two photon microscope controlled with ScanImage<sup>39</sup>. Fluorescence was detected using

a photomultiplier tube (H7422PA-40, Hamamatsu). We used an Olympus  $\times 40$  objective (LUMPlanFI/IR, numerical aperture 0.8) and typically adjusted the power to below 20 mW at the back aperture of the objective. We imaged at a frame rate of 6.7 Hz. Focal planes were selected based on the anatomy and visual responses of microglomeruli. We focused on microglomeruli because this approach allowed us to distinguish different neurons in a labelled population (such distinctions are difficult to make in the ellipsoid body, where axons of different neurons arborize in close proximity).

**Visual stimulation: LED arena.** Visual stimuli were presented using a curved visual display<sup>40</sup> that was covered with a colour filter to prevent cross-talk between fluorescence detection and visual stimulation as described<sup>14</sup>. Additionally, to avoid reflections of stimuli from the curved surface of the display, we covered the display with a diffuser (tracing paper). Under such low-contrast conditions, we found that the signal-to-noise ratio (SNR) of calcium responses when stimulated by dark on bright-background stimuli (for example, dark bar on bright surround) was low. This motivated our preference for bright on dark-background stimuli, which produced higher SNR calcium responses (comparison in Extended Data Fig. 5).

For receptive field measurements the display spanned 270° in azimuth and 120° in elevation. The top left and top right corners of the display (three square panels in each corner, each with a size of 30° by 30°) could not be seen by the fly (because they were occluded by the fly holder) and were excluded from the display. For behavioural experiments, a similar but smaller display was used, spanning 210° in azimuth and 90° in elevation.

**Visual stimulation: flashing dots for receptive field mapping.** Excitatory receptive fields for *c232-GAL4* and *EB1-GAL4* flies were measured with stationary bright square dots (7.5° by 7.5°) appearing randomly in the visual field of the fly for 1 s followed by a dark period of 1 s. The measurements were repeated until the entire visual field covered by the display was stimulated once. The display was sampled with a spatial resolution of the stimulus (7.5°). Measurements were performed in 8 blocks of 140 s each, presenting a total of 468 stimuli covering the entire display.

**Visual stimulation: horizontally moving bars for receptive field mapping.** We used bright bars (15° in elevation, 7.5° in azimuth) that moved at a speed of 30° s<sup>-1</sup> left and right (in azimuthal direction) in the fly's field of view.

Two repetitions with stimulation in one direction were followed by two repetitions of movement in the opposite direction. To map responses over the entire visual field spanned by the display, the stimulus was shifted in steps of 15° in elevation from the top-most part of the visual display towards the bottom part and the left and right moving stimulation was repeated in each row.

**Visual stimulation: white noise.** For white-noise stimulation<sup>37,41</sup> we subdivided the display into squares of 11.25° by 11.25°. Each stimulus frame consisted of 20 randomly selected bright squares whilst the remaining squares remained dark (in a few trials for one fly we used 30 squares). The stimulus appeared for one second followed by a dark period of one second. We presented 60 frames of random stimuli in a trial of 140 s. For *c232-GAL4* flies we used an average of 33  $\pm$  20 trials in 7 flies (59, 61, 29, 21, 21, 14 and 26 trials, respectively). For *EB1-GAL4* flies we used an average of 41  $\pm$  13 trials in 6 flies, (38, 25, 39, 63, 35 and 47 trials, respectively). The colour scale (Fig. 2, Extended Data Fig. 8) was adjusted to give equal weight to excitatory and inhibitory subfields (mean set to 0).

To validate white-noise-based responses, we stimulated the fly with bright bars of 56.25 degrees by 18.75 degrees oriented either vertically, horizontally or rotated by 45 or -45 degrees with respect to the vertical direction. To stimulate the receptive fields of all microglomeruli in an unbiased way, the bar stimuli were presented at a random position on the display and the display was sampled with a resolution of 11.25° in both elevation and azimuth.

**Visual stimulation for measuring orientation tuning curves.** For orientation tuning curve measurement we used a stimulus that extended across the entire display, from 120° in azimuth for vertical orientation (90°) to 270° azimuth for horizontal orientation. The width of the bar was 15° and it was moved at a speed of 75° s<sup>-1</sup>. The angle was changed incrementally in steps of 11.25° starting with either horizontally or vertically oriented bars for one direction of movement, and then repeated in the opposite direction of movement. To obtain tuning curves with dark stimuli we inverted the contrast of bright and dark and removed the diffuser.

**Visual stimulation during walking and flying experiments.** For walking and flying experiments we used a vertical bright bar spanning 90° in elevation and 15° in azimuth. The bar moved horizontally at a velocity of 15° s<sup>-1</sup>. The bar stayed stationary for 10 s after moving in one direction for 17 s and then resumed moving in the opposite direction for 17 s.

**Data analysis.** All data analysis was performed in MATLAB (MathWorks). All errors and error bars shown are s.d. All *P*-values shown are based on *t*-tests, unless otherwise noted.



**Frame alignment and movement correction.** Data recorded using two-photon imaging were aligned in the  $xy$  plane on a frame-by-frame basis. Data were thresholded to distinguish arborizations in the lateral triangle from background. All above-threshold pixels were set to the same value. Frames were aligned by cross-correlating each thresholded frame to a single frame at the beginning of the measurement. Multiple trials were aligned by cross-correlation of trial-averaged frames thresholded as above.

**Calculation of fluorescence changes.** The baseline for calculating  $\Delta F/F$  was selected by averaging over the 10% of frames with lowest intensity in each trial or by using the baseline at the beginning of the experiment. Due to the low baseline intensity of GCaMP6, background fluorescence was not subtracted in measurements with GCaMP6. Calcium traces recorded from behaving flies were smoothed with a third order Savitzky-Golay filter over 7 frames for comparisons with behavioural data.

**Regions of interest selection.** For receptive field measurements in stationary flies, regions of interest (ROIs) corresponding to microglomeruli were selected manually in videos of  $\Delta F/F$ . Overlapping parts of ROIs were excluded from further analysis.

For experiments in behaving flies, regions of interest were selected using visually supervised k-means clustering (Extended Data Fig. 10a–d). We used a subset of three trials (2,820 frames) for clustering-based selection of ROIs. We then used correlation-based k-means clustering between the calcium traces in all thresholded pixels. The number of clusters was selected based on an estimate of the number of microglomeruli. If not all microglomeruli could be separated into different clusters, the number of clusters was increased in a second run. We then set a threshold to remove clusters that were smaller than a certain number of pixels (60). We additionally removed clusters that had an average cross-correlation value lower than a threshold (0.2). We further split anatomically disconnected regions of the same cluster and again removed those parts that were smaller than a size threshold (30 pixels). In a final check, the remaining ROIs were overlaid with the frames of the calcium video that showed the largest response in this ROI, and ROIs that did not correspond to microglomeruli were removed manually.

**Receptive field mapping.** Receptive fields (RFs) were smoothed using a Gaussian filter (4 pixels with a standard deviation of 1 pixel by default; 5 pixels with a standard deviation of 4 pixels for white noise measurements).

For RF measurements (for example, in Fig. 1), we combined responses to left- and right-moving stimuli (each averaged over two trials). The onset of the calcium response was correlated with the bar's movement into the RF. Due to the size of the RF, a bar moving from the right side of the fly towards the left side entered the RF from the right side and induced a calcium onset starting at the right side of the (excitatory part of the) RF. Similarly, a bar entering the receptive field from the left side induced a calcium onset starting at the left side of the RF. Due to the faster on-response than off-response of GCaMP, the onset of the RF was better defined than the offset. To find the centre of the excitatory RF measured with moving bars, we defined the RF centre as the weighted centroid of the average of responses to thresholded left- and right-moving stimuli (Extended Data Fig. 1b). This was equivalent to delimiting the RF by its calcium onset response. This procedure made the location of computed RF centres invariant to the kinetics of the calcium indicator.

We used the following parameters to characterize the excitatory parts of RFs measured with single stationary dots (measured at 50% of the maximum  $\Delta F/F$  response of each RF) (see Extended Data Fig. 7c): area, major axis (of an ellipse that has the same normalized second central moments as the region as determined with the MATLAB 'regionprops' function), minor axis, eccentricity, orientation (the angle between the horizontal  $x$  axis and the major axis of the ellipse, 0 corresponding to horizontal orientation), retinotopic correlation (the correlation coefficient between the centre of the RF, determined as the weighted centroid of the RF area and the centre of the corresponding ROI).

**Receptive field display.** To display microglomerular RFs, we coloured light-emitting diode (LED) arena positions in proportion to the  $\Delta F/F$  response elicited by the stimulus presented in that position. For moving stimuli, the calcium response in each pixel was determined as the average over all response frames that were recorded while the stimulus was at that position. Calcium responses were interpolated to account for mismatches between frame rates and movement of stimuli. For stimulation with stationary stimuli, the  $\Delta F/F$  values shown in RF plots correspond to the maximum  $\Delta F/F$  during the stimulation period. To prevent cross-talk between sequentially presented stimulus frames due to the slow decay of the calcium response (which extended beyond stimulus presentation) we only considered the calcium increase and not the calcium decay in assigning  $\Delta F/F$  values to stimulus frames.

**Retinotopy.** To assess retinotopy, we calculated the correlation coefficient between microglomerulus centres and centres of their RFs. This was compared to the correlation coefficient obtained after randomly shuffling correspondences.

Additionally, we performed principal component analysis (PCA) on the  $(x, y)$  values of RF centres. The first PC gave us the direction of maximum variation of RF centres, which we consider to be the primary retinotopic axis (in the fly's visual field) for the RF population.

**Orientation tuning.** To measure orientation tuning (Fig. 2), single  $15^\circ$ -wide bars spanning the entire visual display were presented, and their orientation changed in steps of  $11.25^\circ$ . Responses from multiple trials were then averaged, and tuning curves were fit using the sum of two circular Gaussians<sup>42,43</sup>:

$$a_1 \cdot e^{-k_1 (\cos(\theta - \theta_1) - 1)} + a_2 \cdot e^{-k_2 (\cos(\theta - \theta_2) - 1)}$$

in which  $a_1$  and  $a_2$  are amplitudes,  $\theta_1$  and  $\theta_2$  are the maximum angles and  $k_1$  and  $k_2$  are width parameters.

The preferred orientation for each microglomerulus was the maximum of its fitted tuning curve.

**Orientation selectivity.** The orientation selectivity index (OSI)<sup>44</sup>, was computed as the difference between the response in the preferred direction,  $\Delta F/F_{\max}$ , and the direction orthogonal to it (preferred direction  $\pm 90^\circ$ ),  $\Delta F/F_{\text{ortho}}$ , normalized by the sum of the responses in the two directions:

$$\text{OSI} = \frac{\Delta F/F_{\max} - \Delta F/F_{\text{ortho}}}{\Delta F/F_{\max} + \Delta F/F_{\text{ortho}}}$$

**Direction selectivity.** The direction selectivity index<sup>42</sup> (DI) was calculated as the difference between responses in the preferred direction,  $\Delta F/F_{\max}$ , and anti-preferred direction (preferred direction  $+180^\circ$ ),  $\Delta F/F_{\text{opposite}}$ , normalized by their sum:

$$\text{DI} = \frac{\Delta F/F_{\max} - \Delta F/F_{\text{opposite}}}{\Delta F/F_{\max} + \Delta F/F_{\text{opposite}}}$$

The sign of the DI was defined as positive for front-to-back movement and negative for back-to-front movement.

**White-noise-based receptive fields and response predictions.** RFs were reconstructed by thresholding calcium traces at 30%  $\Delta F/F$  and averaging over all frames that induced a response larger than this threshold, weighted by peak calcium response. Only the rising slope of the calcium response was considered. The displayed RFs are the weighted averages of stimulus frames and the mean value (background) is subtracted. Max is the maximum of the weighted average after subtracting the mean value, and min is the minimum of this average. The mean value is set to 0.

To predict responses to oriented bars (Fig. 2), we multiplied the mean-subtracted white-noise-based RF with stimulus values in each pixel and summed over all pixels. We convolved the result with a calcium response function. The amplitude and time constants of the calcium response function were fitted using responses to bars of one orientation and then used in the prediction of responses to the remaining three orientations.

The calcium response function,  $crf$ , used in predicting responses to oriented stimuli and in the analysis of calcium responses recorded during behaviour (see below) is given by:

$$crf \sim (1 - (e^{-\frac{t-t_0}{t_{\text{on}}}})) \cdot (e^{-\frac{t-t_0}{t_{\text{off}}}})$$

where  $t_0$  is the onset time and  $t_{\text{on}}$  and  $t_{\text{off}}$  are the rise and decay times of the calcium indicator.

**Walking behaviour analysis.** Ball movement was recorded with a sampling rate of 4 kHz and velocities were calculated with a time base of 250 ms. Velocities were then averaged over all velocity values in each two-photon imaging frame. Average velocities were calculated by detecting epochs in which the fly was moving and then averaged over this period. Owing to strong walking activity in behaving animals, there were few recordings that allowed us to compare microglomerular responses in walking and stationary conditions (Extended Data Fig. 10e–n).

**Flight behaviour analysis.** The wing angle was detected by first manually identifying the wing hinge in both wings. The wing was then detected by first subtracting the background recorded while the fly was not flying, smoothing with a mean filter, and setting a threshold for detecting the wings in two ROIs surrounding the wings. The wing angle was then defined as the angle between the wing hinge and the tip of the wing. Flight was typically intermittent (Extended Data Fig. 10o, p).

**Model for fitting responses of visual microglomeruli during behaviour.** For describing the calcium responses of visual microglomeruli during walking behaviour we fitted the responses with a model consisting of a single excitatory and inhibitory Gaussian function. Model fitting was only performed for microglomeruli that responded with a single peak during one passage of the stimulus on the display, not for bilateral receptive fields. The initial position of each Gaussian, the (common) standard deviation and the amplitude were used as fit parameters.

Since the inhibitory responses could not be directly observed in our calcium signals, we set all negative values to zero.

$$rf \sim (e^{-(x-x_0)^2/(2\sigma_x^2)} - e^{-(x-x_1)^2/(2\sigma_x^2)})$$

$rf = 0$  if  $rf < 0$ ;  $x_0$  and  $x_1$  are the location of the excitatory and inhibitory receptive fields, respectively, and  $\sigma_x$  is the width of the receptive field. The calcium signal  $s$  ( $\Delta F/F$ ) was then modelled as the convolution of the receptive field with the calcium response function,  $crf$ , and a constant offset,  $bg$ :

$$s \sim crf \otimes rf + bg$$

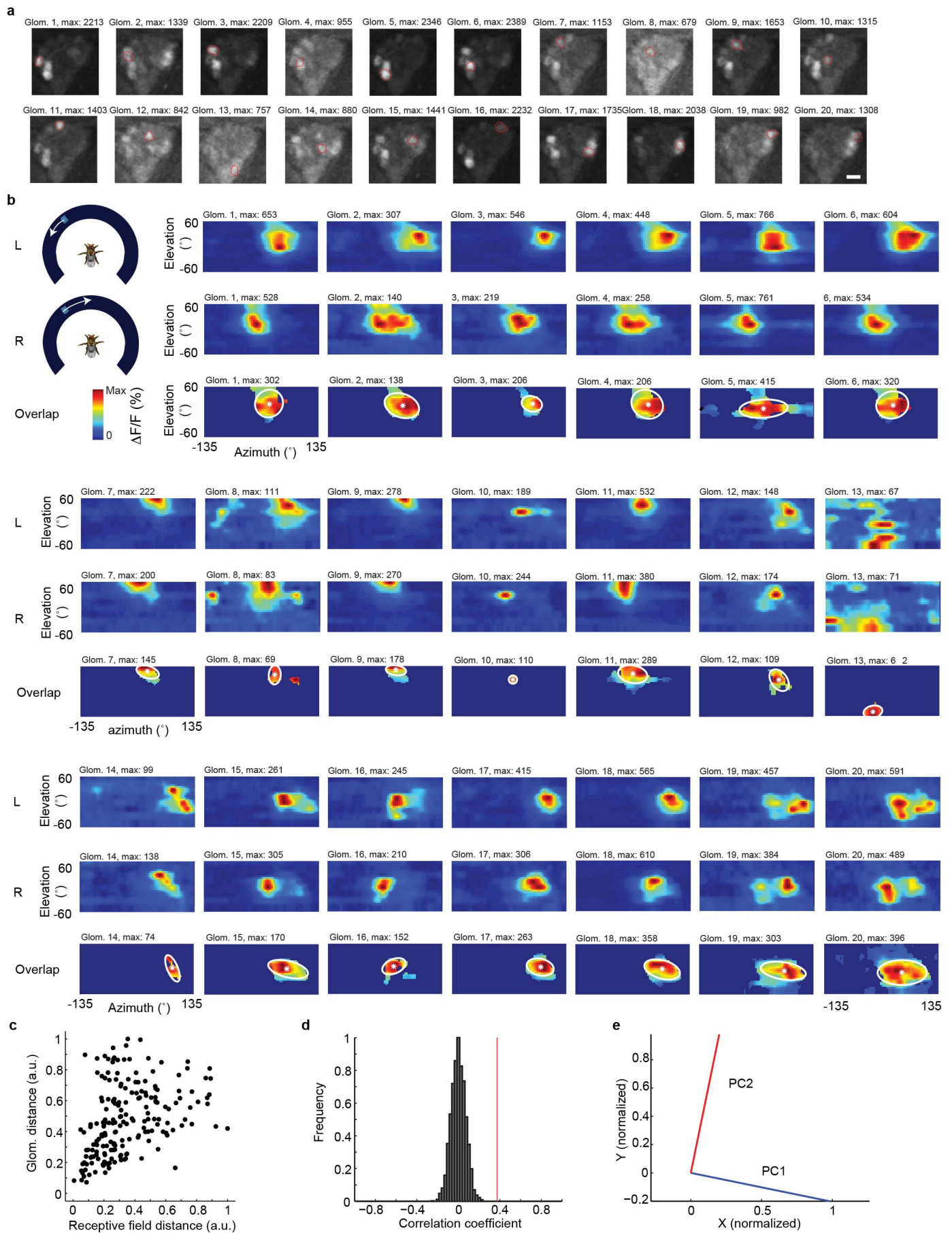
Because the response in flying flies depended on the state of the fly, we used a model for the receptive field that additionally included the left and right wing angles as parameters:

$$rf_{\text{flying}} \sim \text{behaviour} (e^{-(x-x_0)^2/(2\sigma_x^2)} - e^{-(x-x_1)^2/(2\sigma_x^2)})$$

$rf = 0$  if  $rf < 0$ ;  $\text{behaviour} = a \cdot lw + b \cdot rw + c$  depends linearly, with the parameters  $a$ ,  $b$  and  $c$ , on the left wing angle  $lw$  and the right wing angle  $rw$ .

31. Zhang, Y. Q., Rodesch, C. K. & Broadie, K. Living synaptic vesicle marker: synaptotagmin-GFP. *Genesis* **34**, 142–145 (2002).
32. Pfeiffer, B. D. *et al.* Refinement of tools for targeted gene expression in *Drosophila*. *Genetics* **186**, 735–755 (2010).

33. Nicolai, L. J. J. *et al.* Genetically encoded dendritic marker sheds light on neuronal connectivity in *Drosophila*. *Proc. Natl Acad. Sci. USA* **107**, 20553–20558 (2010).
34. Pfeiffer, B. D., Truman, J. W. & Rubin, G. M. Using translational enhancers to increase transgene expression in *Drosophila*. *Proc. Natl Acad. Sci. USA* **109**, 6626–6631 (2012).
35. Struhl, G. & Basler, K. Organizing activity of wingless protein in *Drosophila*. *Cell* **72**, 527–540 (1993).
36. Peng, H. *et al.* BrainAligner: 3D registration atlases of *Drosophila* brains. *Nature Methods* **8**, 493–500 (2011).
37. Weber, F., Machens, C. K. & Borst, A. Spatiotemporal response properties of optic-flow processing neurons. *Neuron* **67**, 629–642 (2010).
38. Suver, M. P., Mamiya, A. & Dickinson, M. H. Octopamine neurons mediate flight-induced modulation of visual processing in *Drosophila*. *Curr. Biol.* **22**, 2294–2302 (2012).
39. Pologruto, T. A., Sabatini, B. L. & Svoboda, K. ScanImage: flexible software for operating laser scanning microscopes. *Biomed. Eng. Online* **2**, 13 (2003).
40. Reiser, M. B. & Dickinson, M. H. A modular display system for insect behavioral neuroscience. *J. Neurosci. Methods* **167**, 127–139 (2008).
41. Chichilnisky, E. J. A simple white noise analysis of neuronal light responses. *Network* **12**, 199–213 (2001).
42. Niell, C. M. & Stryker, M. P. Highly selective receptive fields in mouse visual cortex. *J. Neurosci.* **28**, 7520–7536 (2008).
43. Swindale, N. V. Orientation tuning curves: empirical description and estimation of parameters. *Biol. Cybern.* **78**, 45–56 (1998).
44. Swindale, N. V., Grinvald, A. & Shmuel, A. The spatial pattern of response magnitude and selectivity for orientation and direction in cat visual cortex. *Cereb. Cortex* **13**, 225–238 (2003).

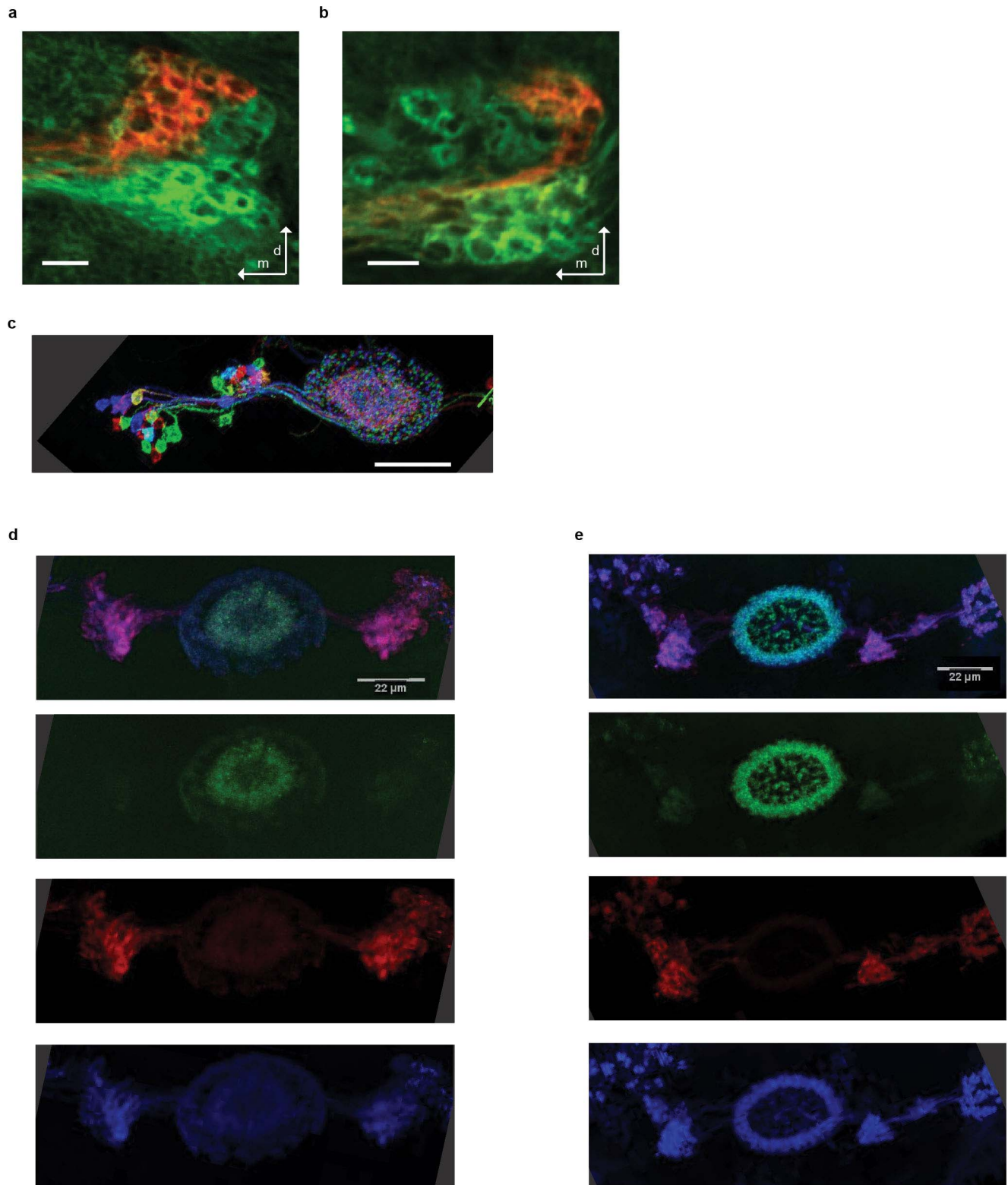




**Extended Data Figure 1 | Pan-neuronal receptive fields for LTR**

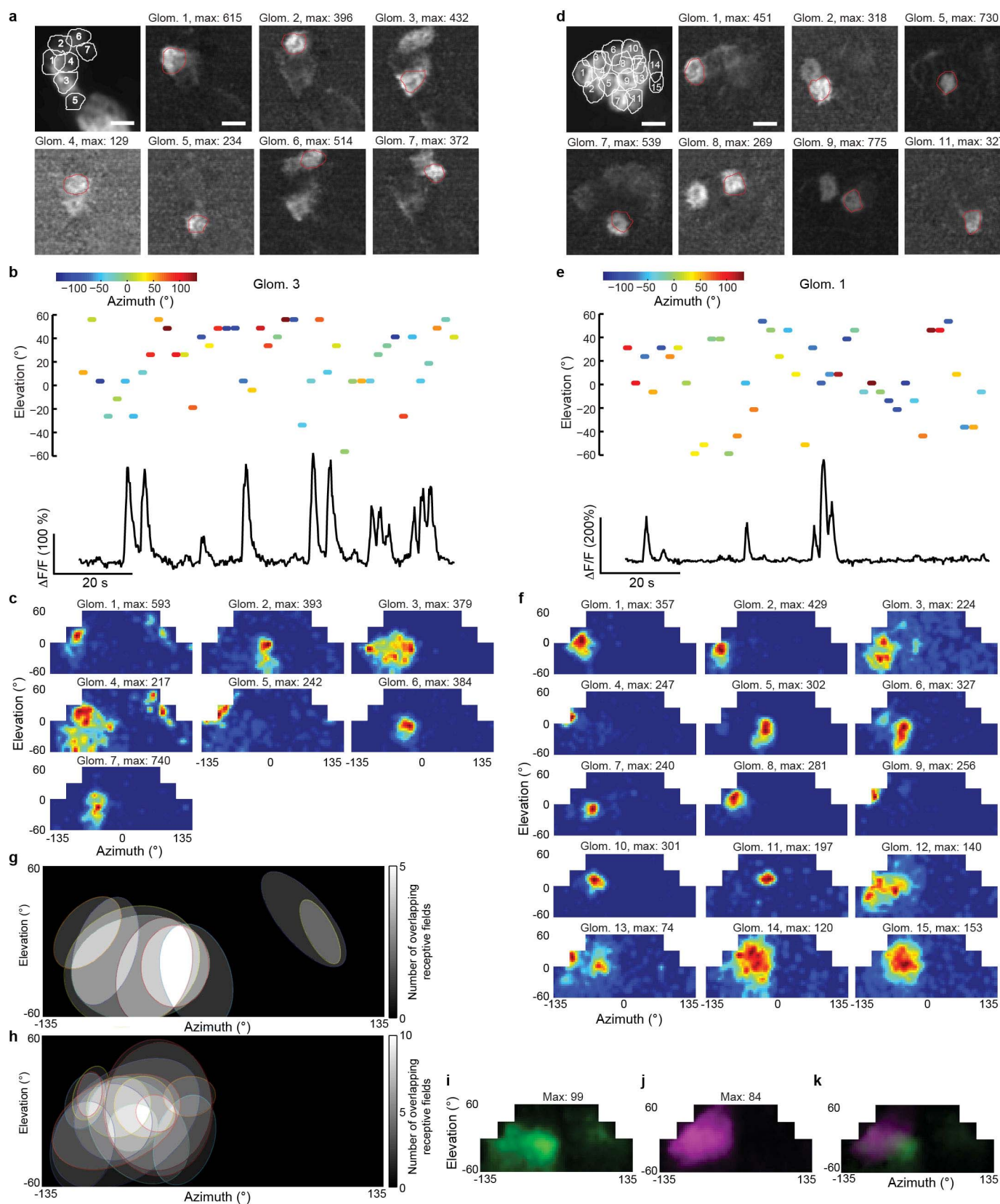
**microglomeruli.** **a**, Selected two-photon imaging frames showing calcium response in the LTR with pan-neuronal expression of GCaMP5. Highlighted in red are the microglomeruli that were selected for receptive field reconstruction. Scale bar, 5  $\mu\text{m}$ . **b**, Schematic of visual stimulation used, receptive fields (two-trial averages) measured with left- and right-moving visual stimuli and their intersection labelled at full width at half maximum (FWHM) intensity (white

ellipse) with the weighted centroid (white asterisk) of the distribution. **c**, Centre of each receptive field plotted against the centre of its corresponding microglomerulus. **d**, Correlation coefficient of data in **d** ( $r = 0.39$ ,  $P = 0$ ) and correlation coefficient of same data after randomly permuting the microglomeruli ( $n = 3,000$  permutations, mean =  $(4.6 \pm 0.076) \times 10^{-4}$ ). **e**, Principal component axes of receptive field centres (see Methods).



**Extended Data Figure 2 | Ring neurons make dendritic projections to LTR microglomeruli and send axonal arbors to the rings of the ellipsoid body.** **a**, A frame of a confocal stack of the LTR labelled with antibody staining against GFP using a pan-neuronal driver line, *R57C10-LexA* and antibody staining against red fluorescent protein (RFP) of *EB1-GAL4*. **d**, dorsal; **m**, medial. **b**, A section of a confocal stack of the LTR labelled with GFP using a pan-neuronal

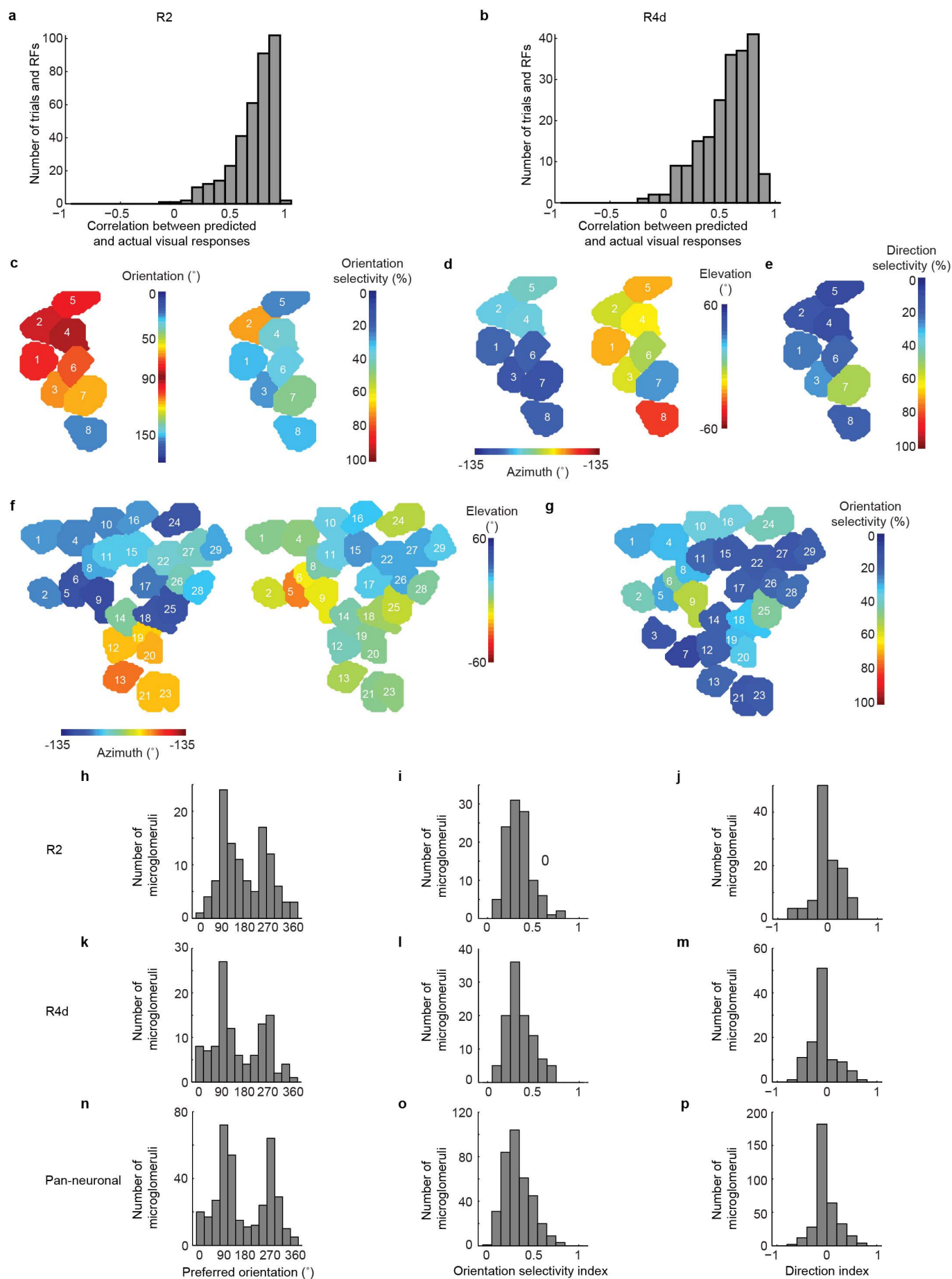
driver line *R57C10-LexA* and RFP labelling of *c232-GAL4*. **c**, Multicolour FLP-out labelling of *c232-GAL4* (see Methods). **d**, Antibody staining of *c232-GAL4* expressing the dendritic marker, DenMark (red) and an axonal label, synaptotagmin-tagged GFP (green). Entire cell is outlined with membrane-tagged HA (blue) (see Methods). Top panel is merge of all colours. **e**, Same as **d** for *EB1-GAL4*.





**Extended Data Figure 3 | Ring neurons respond with strong calcium transients to single stationary bright dots; receptive fields of ring neuron populations are spread non-uniformly across the largely ipsilateral visual field.** **a**, Overlay of all regions of interest of R4d microglomeruli recorded in a fly during visual stimulation over an average of all frames in a calcium video (940 frames), and selected frames of the calcium video showing responses of individual microglomeruli. Scale bar, 5  $\mu\text{m}$ . **b**, An example of the single dot stimulus appearing at the specified azimuth and elevation for 1 s followed by a dark period of 1 s (coloured dot indicates both the stimulus and the subsequent off period) (top).  $\Delta F/F$  for microglomerulus 3 in **a** responding to the stimulus (bottom). **c**, Excitatory receptive fields for all microglomeruli shown in **a**. **d**, Overlay of all regions of interest selected among R2 microglomeruli recorded during visual stimulation over an average of all frames ( $n = 940$ ) in a calcium video, and selected frames of the calcium video showing responses of individual microglomeruli. Scale bar, 5  $\mu\text{m}$ . **e**, Azimuth and elevation of the single bright dot presented for 1 s followed by a 1 s interval without stimulus (top).  $\Delta F/F$  for microglomerulus 1 in **d** responding to the visual stimulus

(bottom). **f**, Excitatory receptive fields for all microglomeruli shown in **d**. **g**, Overlapping receptive fields of R4d neurons. Ellipses have the same normalized second central moments as the receptive fields shown in **c** thresholded at 30%  $\Delta F/F$ . Intensity scale indicates how many receptive fields overlap in a given region. **h**, Same as **g** for R2 neurons shown in **f**. **i–k**, Average receptive fields measured across all flies with single dot stimulation for R4d neurons (green,  $n = 8$  flies, 159 receptive fields) (**i**), for R2 neurons (purple,  $n = 12$  flies, 219 RFs) (**j**) and for both (overlay) (**k**). Recordings from the right lateral triangle were flipped to the contralateral side for averaging. Although a large fraction of R2 and R4d ring neurons we imaged were visually responsive, there were ring neurons that did not respond in our experiments. Such neurons may have receptive fields in areas of the fly's visual field that we did not sample (for example, very lateral areas), be tuned to visual dimensions that we did not explore (such as colour or polarization or complex shapes), or be selectively responsive to other sensory modalities (such as mechanosensation or thermosensation).

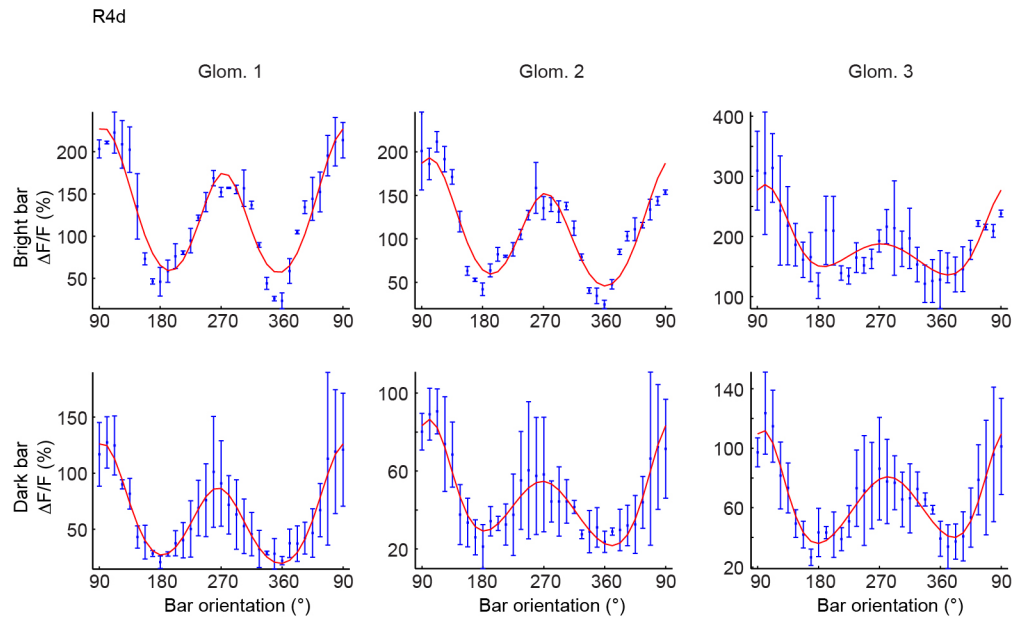
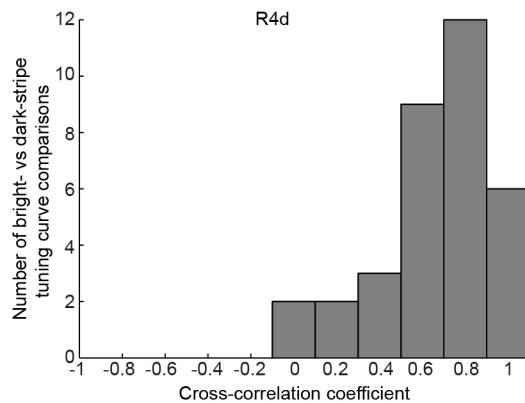


**Extended Data Figure 4 | Ring neuron orientation preference and selectivity, direction selectivity and receptive field centre positions.**

**a**, Correlation coefficients between predicted and measured responses for R2 neurons (histogram over all trials and receptive fields) ( $n = 6$  flies). **b**, Same as **a**, for R4d neurons ( $n = 6$  flies). **c**, Microglomeruli of R4d neurons coloured according to orientation preference and orientation selectivity (six-trial averages). **d**, Azimuth and elevation of the centre of the excitatory receptive fields of the neurons shown in **c** (two-trial averages, measured using moving

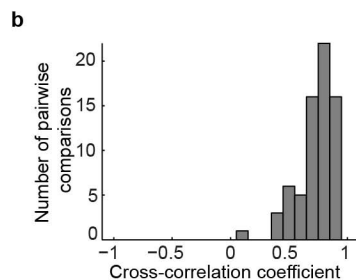
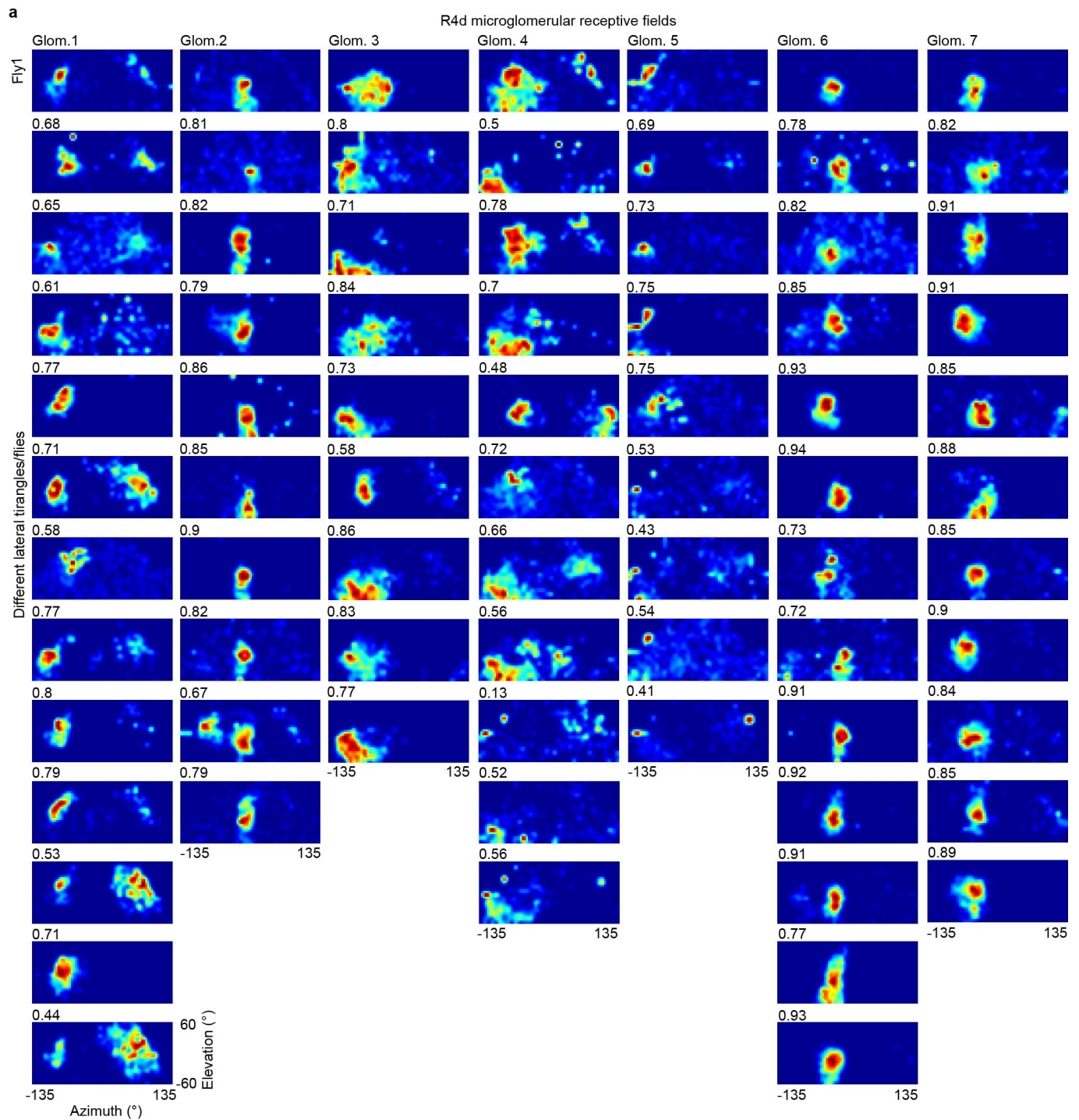
bars as described in Extended Data Fig. 1 and Methods). **e**, Direction selectivity for neurons in **c** (six-trial averages). **f**, **g**, Microglomeruli shown in Fig. 2k, l labelled according to the azimuth and elevation of their excitatory receptive field centre (measured using moving stimuli as in Extended Data Fig. 1) (**f**) and their orientation selectivity (four-trial averages) (**g**). **h–p**, Distributions of preferred orientation, orientation selectivity and direction index for R2 ( $n = 10$  flies), R4d ( $n = 12$ ) and pan-neuronal ( $n = 11$  flies) microglomeruli.



**a****b**

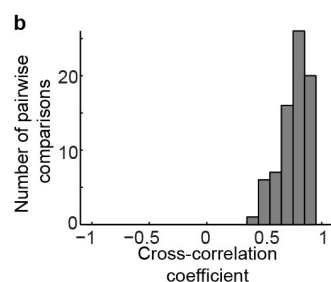
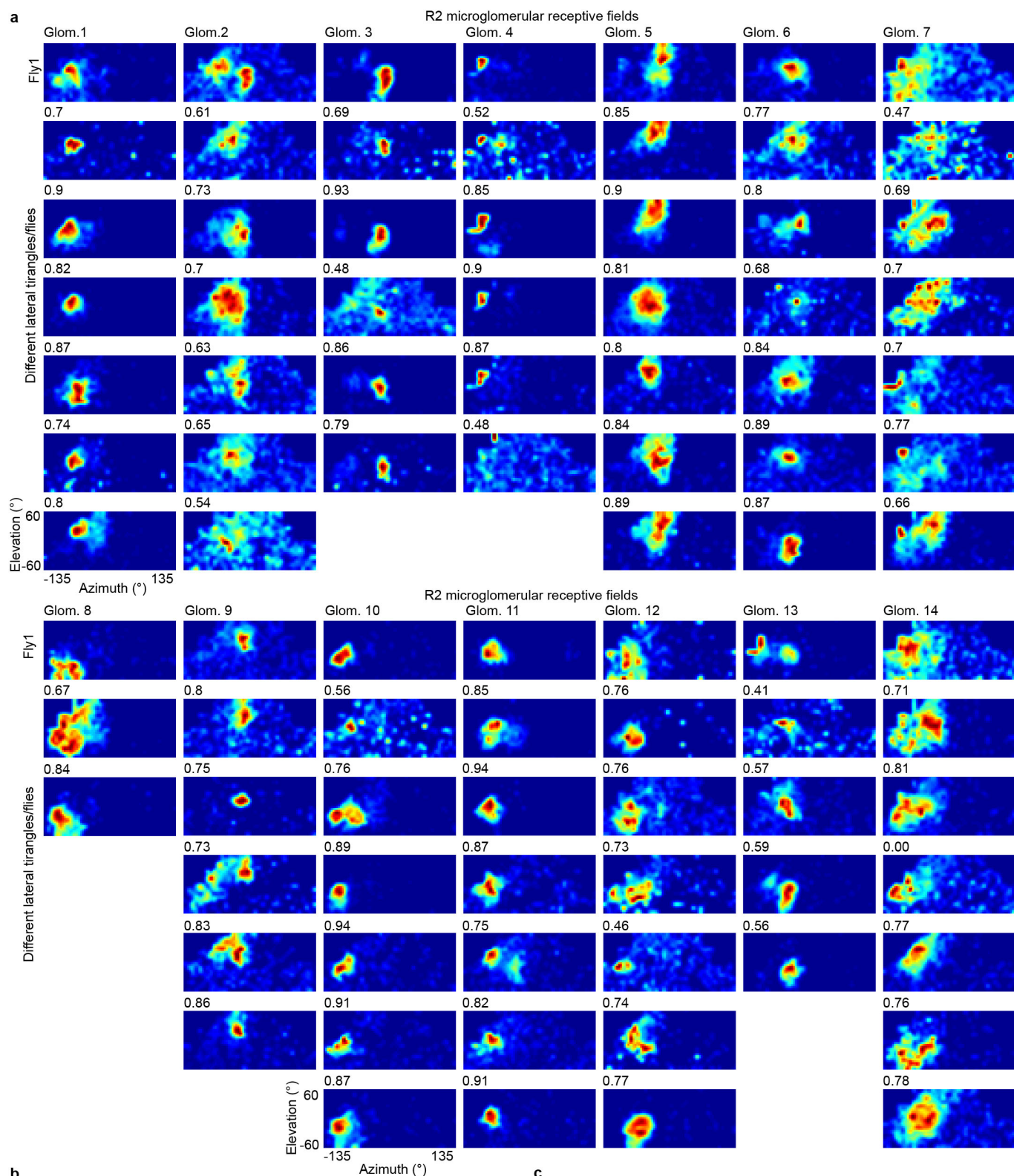
**Extended Data Figure 5 | Comparison of orientation tuning curves measured using dark-on-bright and bright-on-dark bars.** **a**, Orientation tuning curves (GCaMP6s) of three R4d microglomeruli measured using (top row) bright bars on a dark background (as in Fig. 2f) and dark bars on a bright background (exact contrast inversion of bright-bar stimulus). As suggested by

excitatory–inhibitory structure of receptive fields (Fig. 2c, Extended Data Fig. 8), tuning curves measured with the two stimuli are qualitatively similar, but bright bars produce stronger responses (with our low-contrast, filtered LED array display, see Methods). **b**, Fitted tuning curves measured using dark and bright bars are highly correlated (mean =  $0.648 \pm 0.26$ ,  $n = 7$  flies).



**Extended Data Figure 6 | Stereotypy of excitatory receptive fields across flies in R4d neurons. a,** Excitatory receptive fields measured using single dot stimulation in one fly (fly1), and receptive fields recorded in other flies ( $n = 12$ )

aligned by similarity. **b,** Cross-correlation coefficient between fly 1 and all other receptive fields in each column (mean =  $0.74 \pm 0.2$ ).



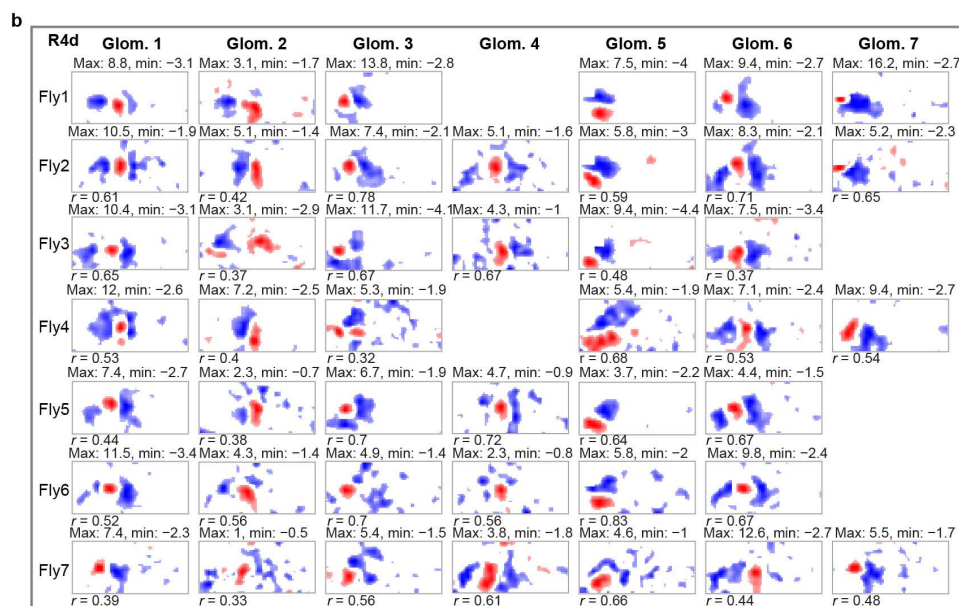
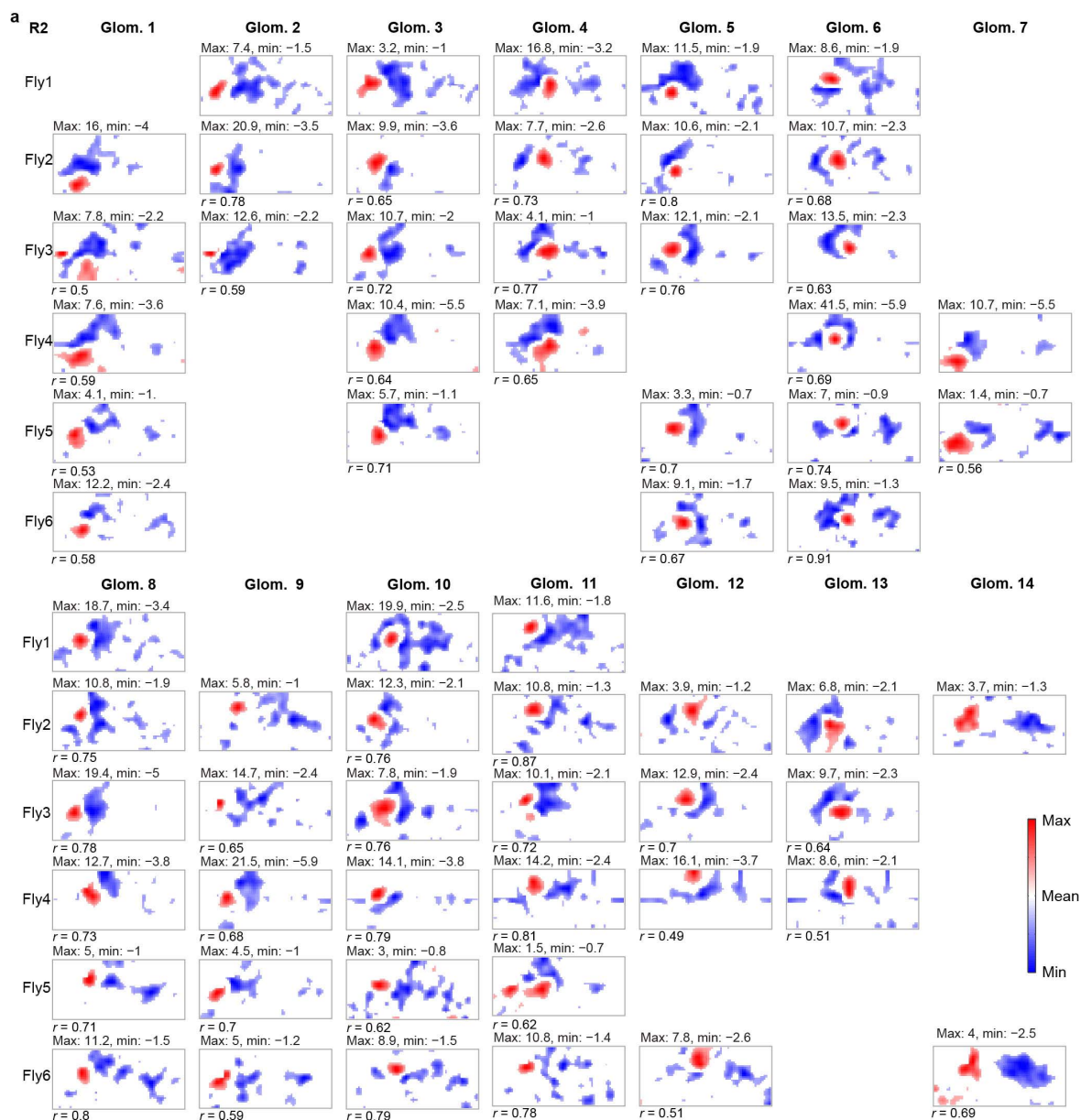
**c**

	R4d	R2
Area [deg <sup>2</sup> ]	994.4 ± 632.2	1,503 ± 1,030
Orientation [deg]	0.0 ± 54.8	-15.0 ± 49.1
Major axis [deg]	47.2 ± 18.6	56.9 ± 22.1
Minor axis [deg]	288 ± 10.6	35.9 ± 13.5
Eccentricity	0.7 ± 0.2	0.7 ± 0.1
Retinotopic correlation	0.18 ± 0.37	0.12 ± 0.2
Retinotopic correlation control	0 ± 0.26	0 ± 0.12



**Extended Data Figure 7 | Stereotypy of excitatory receptive fields across flies in R2 neurons; summary of receptive field properties.** **a**, Same as Extended Data Fig. 6a, measured in R2 neurons in 8 flies. **b**, Cross-correlation coefficient between fly 1 and all other receptive fields in each column for R2 neurons shown in **a** (mean =  $0.75 \pm 0.12$ ). **c**, Table showing characteristics of

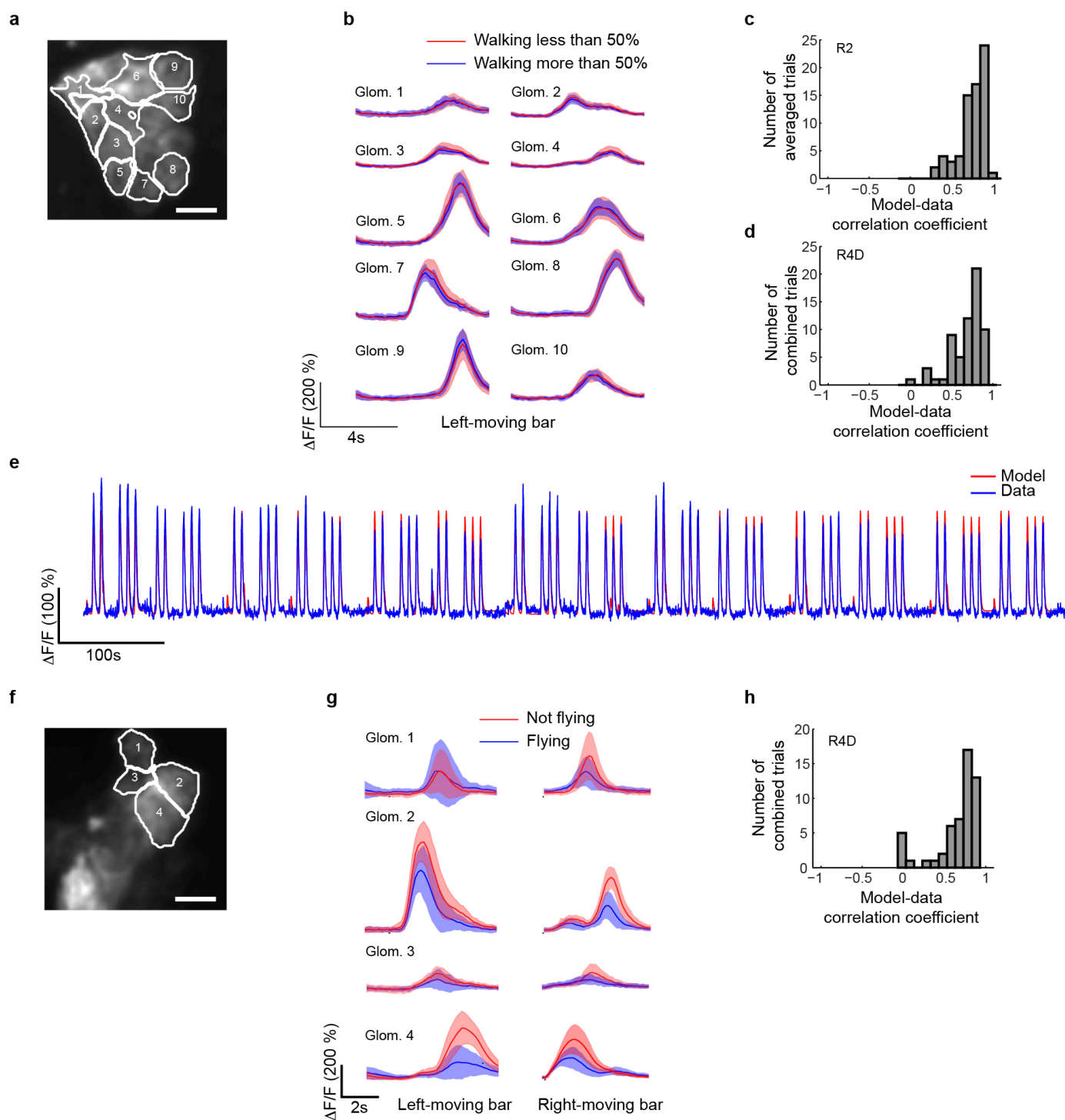
excitatory receptive fields for R4d and R2 ring neurons measured with single dot stimulation calculated at half maximum  $\Delta F/F$ . The retinotopic correlation coefficients are significantly different from the control distributions, with  $P = 0.0283$  for R4d ( $n = 8$  flies, 159 receptive fields) and  $P = 0.0069$  for R2 ( $n = 12$  flies, 219 receptive fields).



**Extended Data Figure 8 | Stereotypy of white-noise-based receptive fields across flies in R2 neurons.** **a**, Receptive fields measured using white-noise stimulation in 6 flies aligned according to similarity. We used an average of  $41 \pm 13$  trials (6 flies, number of trials: 38, 25, 39, 63, 35 and 47). Each trial

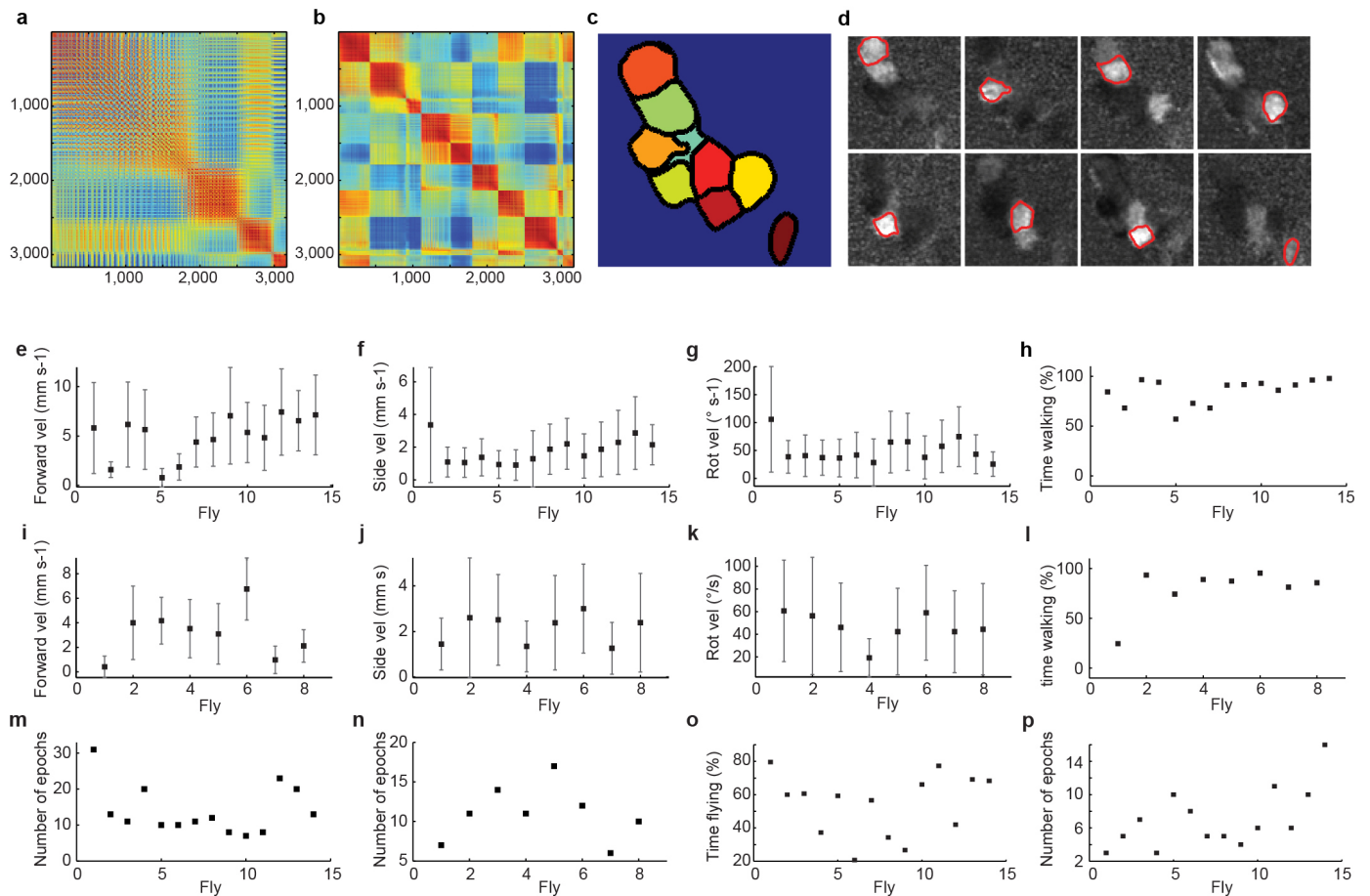
consisted of presentation of 60 stimulus frames (see Methods). **b**, Receptive fields measured using white-noise stimulation in 7 flies aligned by similarity (subset of this data are shown in Fig. 3). We used an average of  $33 \pm 20$  trials (7 flies, number of trials: 59, 61, 29, 21, 21, 14 and 26).





**Extended Data Figure 9 | Ring neuron visual responses in the LTR are diminished during flight but not walking.** **a**, Microglomeruli selected by visually assisted k-means clustering in R2 neurons, a subset of which are shown in Fig. 4b. **b**, Average over  $n = 24$  trials of responses of R2 microglomeruli to a bright bar ( $15^\circ \times 90^\circ$ ) moving at  $30^\circ \text{ s}^{-1}$  shown during a period when the fly is walking more than 50% of the time (blue) and less than 50% of the time (red) during rightward moving stimulation (envelopes show s.d.). **c**, **d**, Correlation coefficient (mean =  $0.72 \pm 0.25$ ) between model (see Methods) and data for R2 microglomeruli recorded in  $n = 8$  flies (**c**), and for R4d microglomeruli ( $n = 14$  flies, mean  $0.7 \pm 0.25$ ) that have unilateral receptive fields (**d**). **e**, Concatenation of seven epochs during walking behaviour

with *EB1-GAL4*, *UAS-GCaMP3* flies during stimulation with a left and right moving bar: data (blue) and model (red) fit (see Methods). **f**, Microglomeruli in R4d selected using visually assisted k-means clustering. **g**, Visual responses during flying and non-flying behaviour during stimulation with a bar moving from left to right (averaged over  $n = 14$  trials for flying,  $n = 14$  trials for non-flying), and during stimulation with a bar moving from right to left ( $n = 17$  for flying and  $n = 11$  trials for non-flying) for the fly shown in Fig. 4e (envelopes show s.d.). **h**, Correlation coefficient between model (see Methods) and data for R4d neurons with a linear dependence of the visual response on both wing angles (mean =  $0.65 \pm 0.27$ ) recorded in  $n = 14$  flies.



**Extended Data Figure 10 | Selection of microglomeruli using visually guided k-means clustering; walking and flight activity during behaving fly experiments.** **a**, Cross-correlation matrix between all pixels above threshold (see Methods for details). **b**, Cross-correlation matrix sorted according to clusters found by k-means clustering. **c**, Resulting regions of interest. **d**, Overlay of resulting regions of interest and selected frames of calcium video. Regions of interest that do not correspond to microglomeruli (based on comparison with the calcium video) are removed. **e–g**, Velocities measured: forward (**e**), side (**f**) and rotational (**g**) for all *c232-GAL4, UAS-GCaMP3* flies recorded during

walking behaviour. **h**, Percentage of walking for all *EB1-GAL4, UAS-GCaMP3* flies recorded. **i–k**, Velocities forward (**i**), sideways (**j**) and rotational (**k**). **l**, Percentage of walking for all *EB1-GAL4, UAS-GCaMP3* flies recorded during walking behaviour. **m**, Number of 130 s epochs recorded for each *c232-GAL4, UAS-GCaMP3* walking fly. **n**, Number of 130 s epochs recorded for each *EB1-GAL4, UAS-GCaMP3* walking fly. **o**, Percentage of time flying during recording for all *c232-GAL4, UAS-GCaMP3* flies recorded during flight behaviour. **p**, Number of 130 s epochs recorded for each *c232-GAL4, UAS-GCaMP3* flying fly.

# SHANK3 and IGF1 restore synaptic deficits in neurons from 22q13 deletion syndrome patients

Aleksandr Shcheglovitov<sup>1</sup>, Olesya Shcheglovitova<sup>1</sup>, Masayuki Yazawa<sup>1</sup>, Thomas Portmann<sup>1</sup>, Rui Shu<sup>1</sup>, Vittorio Sebastiano<sup>2,3</sup>, Anna Krawisz<sup>4</sup>, Wendy Froehlich<sup>4,5</sup>, Jonathan A. Bernstein<sup>4</sup>, Joachim F. Hallmayer<sup>5</sup> & Ricardo E. Dolmetsch<sup>6</sup>

Phelan–McDermid syndrome (PMDS) is a complex neurodevelopmental disorder characterized by global developmental delay, severely impaired speech, intellectual disability, and an increased risk of autism spectrum disorders (ASDs)<sup>1</sup>. PMDS is caused by heterozygous deletions of chromosome 22q13.3. Among the genes in the deleted region is *SHANK3*, which encodes a protein in the postsynaptic density (PSD)<sup>2,3</sup>. Rare mutations in *SHANK3* have been associated with idiopathic ASDs<sup>4–7</sup>, non-syndromic intellectual disability<sup>8</sup>, and schizophrenia<sup>9</sup>. Although *SHANK3* is considered to be the most likely candidate gene for the neurological abnormalities in PMDS patients<sup>10</sup>, the cellular and molecular phenotypes associated with this syndrome in human neurons are unknown. We generated induced pluripotent stem (iPS) cells from individuals with PMDS and autism and used them to produce functional neurons. We show that PMDS neurons have reduced *SHANK3* expression and major defects in excitatory, but not inhibitory, synaptic transmission. Excitatory synaptic transmission in PMDS neurons can be corrected by restoring *SHANK3* expression or by treating neurons with insulin-like growth factor 1 (IGF1). IGF1 treatment promotes formation of mature excitatory synapses that lack *SHANK3* but contain PSD95 and *N*-methyl-D-aspartate (NMDA) receptors with fast deactivation kinetics. Our findings provide direct evidence for a disruption in the ratio of cellular excitation and inhibition in PMDS neurons, and point to a molecular pathway that can be recruited to restore it.

To study cellular and molecular phenotypes associated with PMDS in humans, we generated six iPS cell lines from two patients with PMDS (Supplementary Table 1) that had deletions of approximately 1 megabases in chromosome 22 (Supplementary Fig. 1 and Supplementary Table 2) by using standard reprogramming techniques<sup>11,12</sup>. As controls, we used iPS cell and human embryonic stem (ES) cell lines that were characterized in previous studies<sup>12,13</sup>. Control and PMDS iPS cell lines had typical human ES-cell-like morphologies and expressed endogenous pluripotency marker proteins, but had suppressed the expression of the exogenous reprogramming factors, had normal karyotypes, and formed teratomas *in vivo* (Supplementary Fig. 2 and Supplementary Table 3). We differentiated control and PMDS iPS cells into neurons using a modified monolayer differentiation protocol (Supplementary Fig. 3) and characterized the population of differentiated cells by quantitative PCR with reverse transcription (qRT–PCR) (Supplementary Fig. 4). Similar populations of neuronal cells were found in cultures of control and PMDS neurons. Analysis of gene expression with multiplex single-cell qRT–PCR<sup>13,14</sup> demonstrated that approximately 42% of cells expressed pan-neuronal markers, 15% expressed cortical layer markers, and only 2–5% of the cells expressed the synaptic markers *SHANK1–SHANK3* (Supplementary Fig. 5), indicating that only a small fraction of the cells were synaptically mature neurons. To enrich for mature forebrain neurons, we infected cells with enhanced green fluorescent protein (GFP) under the control of the CaMKII $\alpha$  promoter (CaMKII $\alpha$ -GFP)<sup>15,16</sup> (Fig. 1a). Single-cell profiling

of GFP-expressing cells showed an approximate 12-fold enrichment in cells expressing *SHANK1–SHANK3* (Fig. 1b). These cells also expressed synaptic markers and genes typically found in the cortex and striatum, including *TBR1*, *CUX1*, *SATB2*, *SOX5*, *FOXP1*, *CTIP2* (also known as *BCL11B*) and *FOXP2* (Supplementary Fig. 5) and were stained with antibodies recognizing MAP2 (~99%), TBR1 (~29%), CTIP2 (~66%), SATB2 (~42%), and  $\gamma$ -aminobutyric acid (GABA) or GAD67 (~7.5%) (Fig. 1c–h and Supplementary Fig. 6).

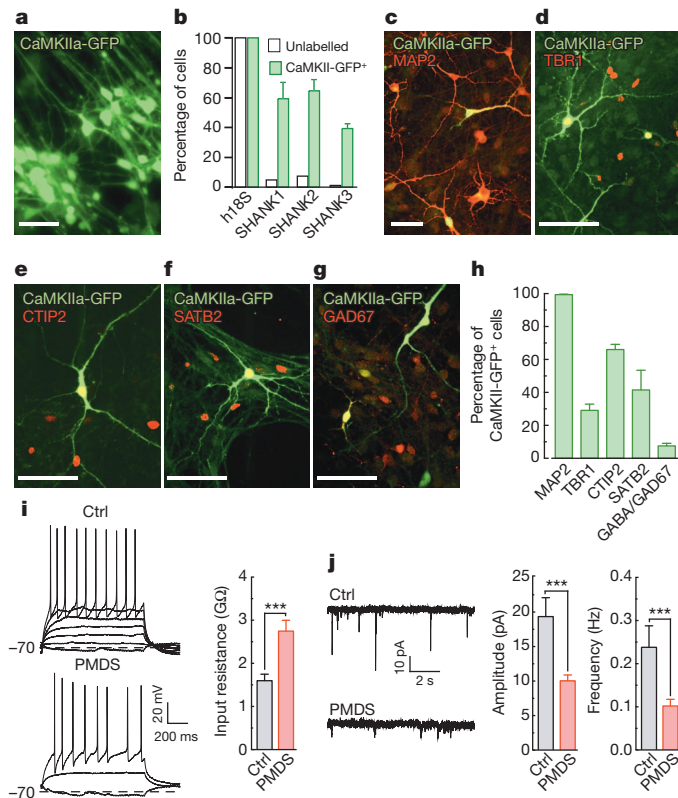
We next investigated the electrophysiological properties of control and PMDS neurons. Control and PMDS neurons expressing GFP had similar action potential characteristics, resting membrane potential and capacitance (Supplementary Table 4). PMDS neurons, however, had an increased input resistance and reduced amplitude and frequency of spontaneous synaptic events (Fig. 1i–j and Supplementary Table 4). A reduction in synaptic transmission in PMDS neurons could arise from differences in the number of neurons present in cultures of PMDS and control neurons. In fact, anti-MAP2 antibody staining revealed that our PMDS iPS cell lines produced slightly fewer neurons than controls (Supplementary Fig. 7). To control for this difference in neuronal density and to study cells of both genotypes under identical experimental conditions, we infected control and PMDS neurons with either CaMKII $\alpha$ -GFP or CaMKII $\alpha$ -mKate2 and grew them together in the same dish (Fig. 2a). Under these conditions, we found that the amplitude and frequency of spontaneous excitatory postsynaptic currents (EPSCs) recorded from PMDS neurons were significantly reduced relative to controls (Fig. 2b, c), indicating that synaptic defects in PMDS neurons are postsynaptic and cell-autonomous. To determine if the changes in spontaneous EPSCs were also reflected in differences in evoked  $\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) and NMDA-mediated synaptic responses, we measured currents induced by local field stimulation at  $-70$  mV (AMPA-EPSCs) and  $+60$  mV (NMDA-EPSCs). Smaller AMPA- and NMDA-EPSCs were measured in PMDS neurons in response to different stimulus intensities (Fig. 2d–f), indicating that both AMPA- and NMDA-mediated synaptic transmissions are significantly impaired.

To determine if inhibitory synaptic transmission was also altered in PMDS neurons, we recorded spontaneous and evoked inhibitory postsynaptic currents (IPSCs) (Fig. 2g–j). There were no differences in IPSCs measured from co-cultured control and PMDS neurons. To determine the direction of the voltage shifts induced by IPSCs at this developmental stage, we measured voltage deflections induced by focal  $\gamma$ -aminobutyric acid application (Supplementary Fig. 8). Most control and PMDS neurons responded with hyperpolarizing shifts in membrane potential. These results indicate that the defects in PMDS neurons are limited to excitatory synaptic transmission and suggest that there is an alteration in the balance of cellular excitation and inhibition in PMDS neurons.

To gain insight into the mechanism underlying the impairment of excitatory synaptic transmission in PMDS neurons, we measured the

<sup>1</sup>Department of Neurobiology, Stanford University, Stanford, California 94305, USA. <sup>2</sup>Department of Obstetrics and Gynecology, Stanford University, Stanford, California 94305, USA. <sup>3</sup>Institute for Stem Cell Biology and Regenerative Medicine, Stanford University, Stanford, California 94305, USA. <sup>4</sup>Department of Pediatrics, Stanford University, Stanford, California 94305, USA. <sup>5</sup>Department of Psychiatry and Behavioral Science, Stanford University, Stanford, California 94305, USA. <sup>6</sup>Novartis Institutes for Biomedical Research, Cambridge, Massachusetts 02139, USA.



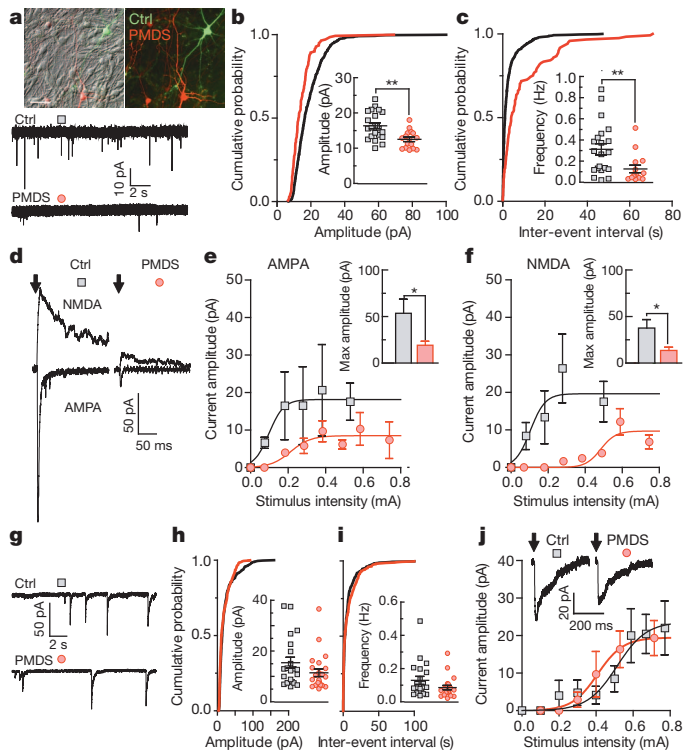


expression levels of AMPA and NMDA receptors. PMDS neurons expressed significantly less GLUA1 and GLUN1 proteins (Fig. 3a) and generated significantly smaller currents in response to the focal application of either AMPA (Fig. 3b) or NMDA (Fig. 3c). In contrast, the amplitude of  $\gamma$ -aminobutyric-acid-evoked currents was similar for both control and PMDS neurons (Supplementary Fig. 9), indicating that PMDS neurons express a decreased number of excitatory neurotransmitter receptors and normal levels of inhibitory receptors.

To determine whether the impairment in excitatory synaptic transmission in PMDS neurons is due to a reduced number of synapses, we immunostained co-cultured control and PMDS neurons with antibodies that recognize the presynaptic protein synapsin 1 (SYN1) and postsynaptic protein HOMER1 (Fig. 3d). We observed a significant decrease in the number of puncta that were SYN1 $^{+}$  (Fig. 3e) and HOMER1 $^{+}$  SYN1 $^{+}$  (Fig. 3f) in PMDS neurons as compared to controls, indicating that PMDS neurons have significantly fewer excitatory synapses than controls.

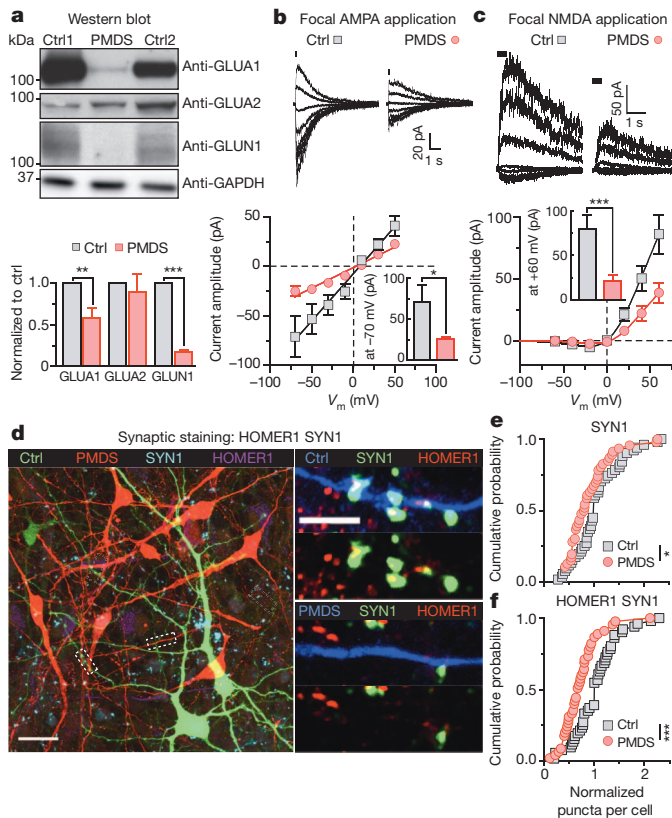
Although SHANK3 has been proposed to be primarily responsible for the neurological phenotype in PMDS patients, the vast majority of PMDS patients have 10–100 additional genes deleted. To determine whether the synaptic phenotypes in PMDS neurons are due to reduced SHANK3 expression, we measured the levels of SHANK3 messenger RNA and protein. Whereas MAP2, SHANK1 and SHANK2 were expressed at similar levels in control and PMDS neurons (Fig. 4a), SHANK3 mRNA was significantly reduced in PMDS (Fig. 4a). Western blot analysis showed a significant reduction in the expression of the longest SHANK3 protein isoform (Fig. 4b), which was expressed primarily in neurons (Supplementary Fig. 10). Immunostaining with anti-SHANK3 antibodies on co-cultured control and PMDS neurons also revealed a significantly reduced level of SHANK3 protein expression in PMDS neurons (Fig. 4c), indicating that loss of a copy of SHANK3 in PMDS significantly downregulates the expression of the protein.

We next investigated whether we could reverse the synaptic phenotype in PMDS neurons by increasing SHANK3 expression. We infected PMDS neurons with a lentivirus containing the longest isoform of SHANK3 fused to EGFP (Fig. 4d and Supplementary Fig. 11). SHANK3 expression in PMDS neurons restored both the amplitude and frequency of spontaneous EPSCs (Fig. 4e–g). It also completely rescued evoked AMPA- and NMDA-EPSCs in 43% of recorded cells, and in



Although SHANK3 has been proposed to be primarily responsible for the neurological phenotype in PMDS patients, the vast majority of PMDS patients have 10–100 additional genes deleted. To determine whether the synaptic phenotypes in PMDS neurons are due to reduced SHANK3 expression, we measured the levels of SHANK3 messenger RNA and protein. Whereas MAP2, SHANK1 and SHANK2 were expressed at similar levels in control and PMDS neurons (Fig. 4a), SHANK3 mRNA was significantly reduced in PMDS (Fig. 4a). Western blot analysis showed a significant reduction in the expression of the longest SHANK3 protein isoform (Fig. 4b), which was expressed primarily in neurons (Supplementary Fig. 10). Immunostaining with anti-SHANK3 antibodies on co-cultured control and PMDS neurons also revealed a significantly reduced level of SHANK3 protein expression in PMDS neurons (Fig. 4c), indicating that loss of a copy of SHANK3 in PMDS significantly downregulates the expression of the protein.

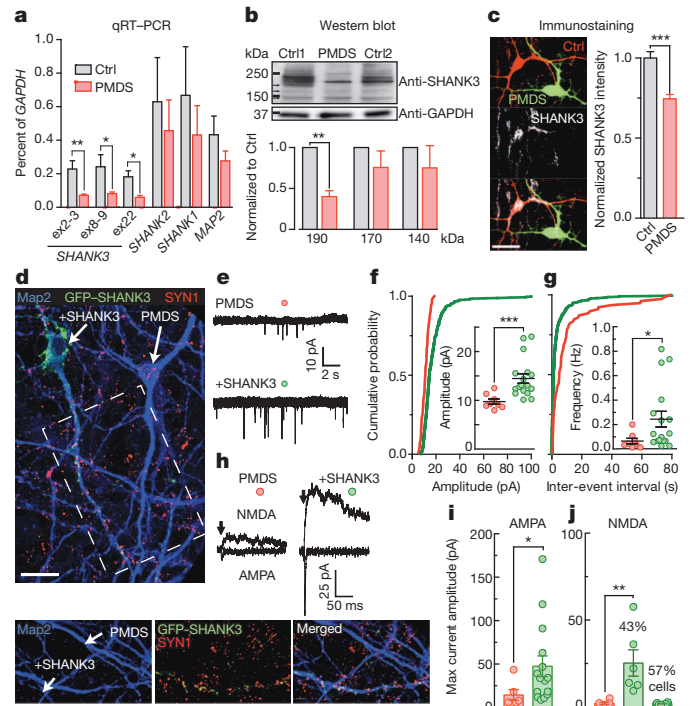
We next investigated whether we could reverse the synaptic phenotype in PMDS neurons by increasing SHANK3 expression. We infected PMDS neurons with a lentivirus containing the longest isoform of SHANK3 fused to EGFP (Fig. 4d and Supplementary Fig. 11). SHANK3 expression in PMDS neurons restored both the amplitude and frequency of spontaneous EPSCs (Fig. 4e–g). It also completely rescued evoked AMPA- and NMDA-EPSCs in 43% of recorded cells, and in



**Figure 3 | PMDS neurons show reduced expression of glutamate receptors and decreased number of synapses.** **a**, Quantification of expression of AMPA and NMDA receptors in control and PMDS neurons using western blot (day *in vitro* 45–50;  $n = 3$ –11 pairs of samples;  $**P < 0.01$ ,  $***P < 0.001$ , one-sample *t*-test). GAPDH was used as a loading control. **b**, **c**, Representative currents (top) and peak current versus voltage curves (bottom) recorded in control and PMDS neurons in response to focal application of 200  $\mu$ M AMPA (measured at a holding potential of  $-70$  to  $+50$  mV,  $\Delta 20$  mV,  $n = 7$  control and 11 PMDS cells) and 100  $\mu$ M NMDA/10  $\mu$ M Gly ( $-60$  to  $+60$  mV,  $\Delta 20$  mV,  $n = 21$  control and 30 PMDS cells). Data presented as means  $\pm$  s.e.m.;  $*P < 0.05$ ,  $***P < 0.001$ , Mann–Whitney test. **d**, Images of co-cultured control (green) and PMDS (red) neurons immunostained for mKate2, GFP, SYN1 and HOMER1. Scale bars, 20 (left) and 5  $\mu$ m (right). **e**, **f**, Cumulative distributions of the number of SYN1 $^{+}$  (**e**) and HOMER1 $^{+}$  SYN1 $^{+}$  (**f**) puncta on co-cultured control and PMDS neurons;  $*P < 0.05$ ,  $***P < 0.001$ , Kolmogorov–Smirnov test.

the remaining cells, it rescued only AMPA-EPSCs (Fig. 4h–j). This indicates that loss of *SHANK3* substantially contributes to the impaired excitatory synaptic transmission detected in PMDS neurons; however, it also suggests that *SHANK3* may act differently on AMPA and NMDA receptors.

We next tested if we could restore synaptic deficits in PMDS neurons pharmacologically, using agents that have been previously reported to increase the level of *SHANK3* expression or to increase synaptic transmission, including trichostatin A (TSA)<sup>17,18</sup>, valproic acid (VPA)<sup>17,19</sup>, nifedipine (Nif)<sup>20</sup>, IGF1<sup>21–23</sup> and IGF2<sup>24</sup>. We treated control and PMDS neurons over the course of 2 to 20 days and measured the number of HOMER1 $^{+}$  SYN1 $^{+}$  puncta. Only IGF1 increased the number of SYN1 $^{+}$  HOMER1 $^{+}$  puncta to the levels observed in control neurons in the same culture (Fig. 5a–b). We also found that IGF1 increased the average amplitude and frequency of spontaneous EPSCs in PMDS neurons to control levels (Fig. 5c), but had no effect on spontaneous IPSCs (Supplementary Fig. 12). IGF1 also restored the amplitude of evoked AMPA- and NMDA-EPSCs (Fig. 5d) and reduced the input resistance of PMDS neurons to that observed in control neurons (Supplementary Fig. 13). Finally, IGF1 completely restored the

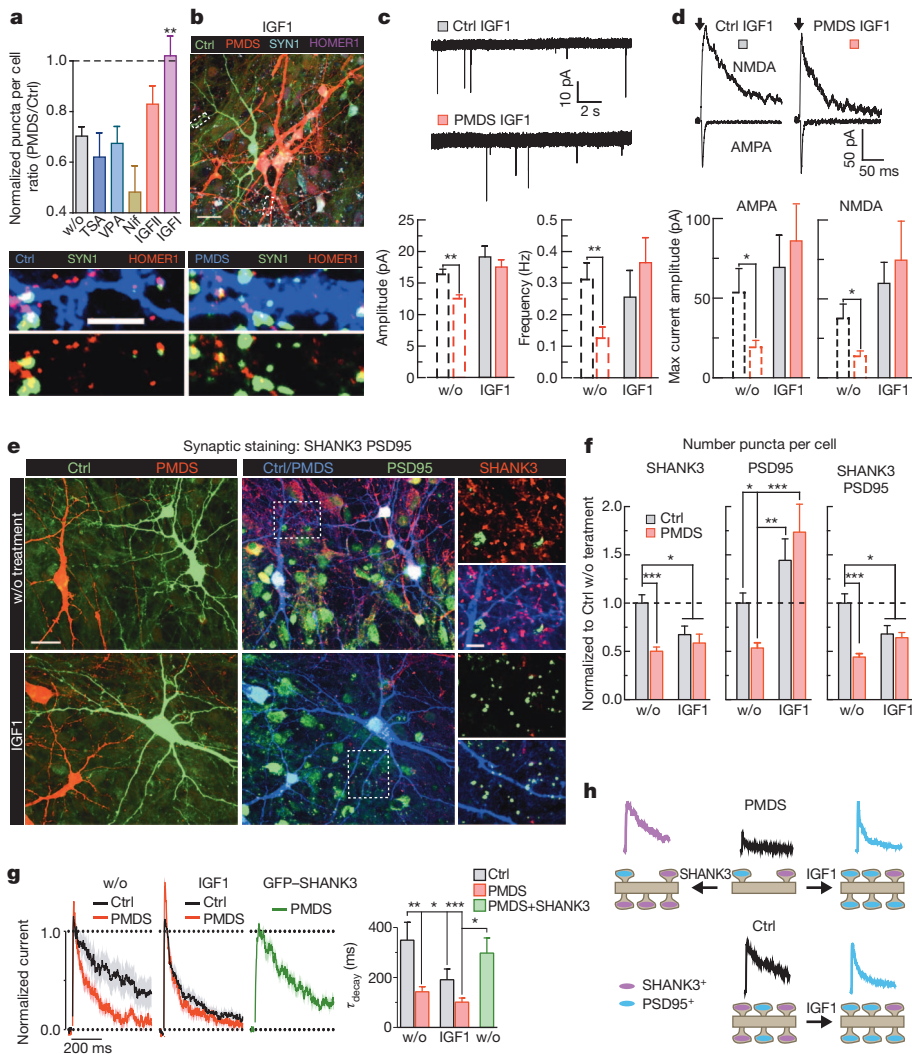


**Figure 4 | Reduced *SHANK3* expression contributes to synaptic defects in PMDS neurons.** **a**, Relative expression of *SHANK1*, *SHANK2*, *SHANK3* and *MAP2* mRNAs in iPS cell-derived neurons (day *in vitro* 30–35;  $n = 5$  control and 6 PMDS biological replicates;  $*P < 0.05$ ,  $**P < 0.01$ , Student's *t*-test). ex, exon. **b**, Quantification of *SHANK3* expression in control and PMDS neurons as assessed by western blot (day *in vitro* 45–50;  $n = 4$  pairs of samples;  $***P < 0.01$ , one-sample *t*-test). GAPDH was used as a loading control. **c**, Images (left) and quantification (right) of *SHANK3* expression in co-cultured control (red) and PMDS (green) neurons (scale bar, 20  $\mu$ m). **d**, Images of PMDS neurons infected with GFP-tagged *SHANK3* lentiviruses and immunostained with antibodies against MAP2, GFP and SYN1 (scale bars, 20 (top) and 5 (bottom)  $\mu$ m). **e**, Traces of spontaneous EPSCs recorded from uninfected and *SHANK3*-infected PMDS neurons at  $-70$  mV. **f**, **g**, Cumulative distribution and quantification of amplitude (**f**) and frequency (**g**) of spontaneous EPSCs ( $n = 8$  uninfected and 17 *SHANK3*-infected cells). **h**, Traces of evoked AMPA- ( $V_h = -70$  mV) and NMDA-EPSCs ( $V_h = +60$  mV). **i**, **j**, Quantification of maximum amplitudes of evoked AMPA- (**i**, measured at peak) and NMDA-EPSCs (**j**, measured 50 ms post stimulus) recorded from uninfected ( $n = 6$ ) and *SHANK3*-infected ( $n = 14$ ) PMDS neurons. Data presented as means  $\pm$  s.e.m.;  $*P < 0.05$ ,  $***P < 0.001$ , Mann–Whitney test.

whole-cell NMDA receptor currents evoked by focal application of NMDA, but failed to fully restore whole-cell AMPA receptor currents elicited by application of AMPA (Supplementary Fig. 14). Together, these results indicate that treatment with IGF1 restores synaptic transmission in PMDS neurons by increasing the number of synaptic AMPA and NMDA receptors.

To gain insight into the mechanism of IGF1 action, we measured *SHANK3* expression in control and PMDS neurons treated with IGF1. Surprisingly, we found that IGF1 caused a marked decrease of *SHANK3* protein expression in the cell body and processes of both control and PMDS neurons (Supplementary Fig. 15). We next examined *SHANK3* expression at synapses by infecting cells with GFP-tagged *SHANK3* and staining neurons with the postsynaptic antibody anti-PSD95 (Supplementary Fig. 16a). We found only 50% overlap between *SHANK3*- and PSD95-containing puncta in human iPS cell-derived neurons. A similar pattern of *SHANK3* and PSD95 expression was detected in the human fetal brain tissue immunostained with anti-*SHANK3* and anti-PSD95 antibodies (Supplementary Fig. 16b). This indicates that in developing neurons there are three populations of excitatory synapses: those that contain *SHANK3* alone, those that contain *SHANK3* and





**Figure 5 | IGF1 treatment restores excitatory synaptic transmission in PMDS neurons.**

**a**, Quantification of the number of HOMER1<sup>+</sup> SYN1<sup>+</sup> puncta in co-cultured neurons after treatment with different agents (w/o:  $n = 51$  control, 43 PMDS cells; TSA:  $n = 11/12$ , VPA:  $n = 25/29$ , Nif:  $n = 9/8$ , IGF2:  $n = 31/40$ , IGF1:  $n = 35/32$ ). w/o, without treatment. **b**, Images of co-cultured control (green) and PMDS (red) neurons after IGF1 treatment. Scale bars, 20 (top) and 5 (bottom)  $\mu\text{m}$ . **c**, Recordings (top) and quantification of amplitude and frequency (bottom) of spontaneous EPSCs recorded in IGF1-treated cultures at  $-70$  mV ( $n = 18$  control and 25 PMDS cells). **d**, Recordings (top) and quantification of amplitudes (bottom) of evoked AMPA- (measured at the peak,  $V_h = -70$  mV) and NMDA-EPSCs (measured 50 ms post-stimulus,  $V_h = +60$  mV) acquired in IGF1-treated cultures ( $n = 14$  control and 21 PMDS cells). Dashed lines represent data originally presented in Fig. 2. **e**, Images of co-cultured control (green) and PMDS (red) neurons in untreated (top) and IGF1-treated (bottom) cultures, immunostained with antibodies against GFP, mKate2, SHANK3 and PSD95. Scale bars, 20 (left) and 5 (right)  $\mu\text{m}$ . **f**, Quantification of the number of SHANK3<sup>+</sup>, PSD95<sup>+</sup> and SHANK3 PSD95<sup>+</sup> puncta in co-cultured control (w/o,  $n = 25$  cells; IGF1, 21 cells) and PMDS (w/o,  $n = 33$  cells; IGF1, 25 cells) neurons. Values were normalized to the mean values acquired from control neurons in untreated cultures. **g**, Mean normalized traces (left) and quantification of the decay kinetics (right) of NMDA-EPSCs recorded from control (w/o,  $n = 9$ ; IGF1, 12) and PMDS (w/o,  $n = 18$ ; IGF1, 18, GFP-SHANK3, 6) neurons. Data presented as means  $\pm$  s.e.m.; \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , one-way analysis of variance with Bonferroni's test. **h**, Cartoon depicting possible mechanisms for the recovery of synaptic transmission in PMDS neurons.

PSD95, and those that contain PSD95 alone. We investigated the effect of IGF1 on these three populations of synapses by staining neurons with anti-SHANK3 and anti-PSD95 antibodies (Fig. 5e). Treatment with IGF1 caused a decrease in the number of puncta expressing SHANK3 alone in control neurons, but had little effect in PMDS neurons, which already had reduced levels of SHANK3 (Fig. 5f). In contrast, IGF1 caused a 45% and 340% increase in the fraction of puncta expressing PSD95 in control and PMDS neurons, respectively, while decreasing the number of puncta containing both markers (Fig. 5f). To determine if synapses that lack SHANK3 are functionally different, we analysed the physiological properties of synapses in PMDS and control neurons. We found that PMDS neurons have a significantly faster rate of decay of NMDA-EPSCs relative to control cells and treatment with IGF1 increased the decay rate of NMDA-EPSCs in control neurons to match the rate of decay in PMDS neurons (Fig. 5g). In contrast, SHANK3 expression in PMDS neurons caused a decrease in the rate of decay of NMDA-EPSCs to match that in untreated control neurons (Fig. 5g). These results are consistent with the idea that IGF1 treatment rescues synaptic transmission in PMDS neurons by causing loss of SHANK3-containing synapses and gain of PSD95-containing synapses, which are characterized by rapidly decaying NMDA-EPSCs (Fig. 5h).

Here, we show that neurons derived from iPS cells from PMDS patients have major defects in excitatory synaptic transmission arising from both a failure to form the correct number of excitatory synapses and a reduction in the expression of glutamate receptors. As no significant impairments in inhibitory synaptic transmission were found,

this resulted in an imbalance between excitation and inhibition at a cellular level. Only some aspects of the phenotype detected in human cells are similar to those found in mouse models of PMDS<sup>25–27</sup>. Why there is a discrepancy is unclear. One potentially important issue is that there are multiple SHANK3 isoforms. PMDS patients generally lack the entire gene on one chromosome, whereas mice typically lack only some of the isoforms and have generally been studied as homozygote knockouts. We also demonstrated that SHANK3 and IGF1 largely restore synaptic deficits in PMDS neurons. Interestingly, IGF1 decreases SHANK3 expression and promotes the formation of a class of synapses containing PSD95 but lacking SHANK3. One possibility is that SHANK3 is required at an early stage of synapse formation, and that IGF1 promotes maturation of synapses by triggering the loss of SHANK3 and recruitment of PSD95. Consistent with this view, SHANK3-containing synapses have slowly decaying NMDA receptor currents that resemble those of GLUN2B receptors that occur early in development, whereas synapses that contain PSD95 but lack SHANK3 have rapidly decaying NMDA receptor currents that resemble the GLUN2A-containing synapses that appear later in development<sup>28</sup>. In summary, we have identified synaptic defects in iPS cell-derived neurons from patients with PMDS and we show that these defects depend on loss of SHANK3 and can be reversed by IGF1.

## METHODS SUMMARY

Primary fibroblasts from consenting patients with PMDS were acquired by skin-punch biopsy following approved SCRO and IRB protocols. iPS cell lines from the fibroblasts were generated using standard reprogramming techniques<sup>11,12</sup>. iPS cells



were differentiated into neurons using a modified monolayer differentiation protocol (Supplementary Fig. 3). Properties of iPS cell-derived neurons were characterized using qRT-PCR, multiplex single-cell qRT-PCR, electrophysiology, immunocytochemistry and western blot. A detailed description of all the procedures, materials and methods used in this study is available in the Methods section. Performed experiments are also summarized in Supplementary Table 3.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 8 November 2011; accepted 30 August 2013.**

**Published online 16 October 2013.**

- Phelan, K. & McDermid, H. E. The 22q13.3 deletion syndrome (Phelan-McDermid syndrome). *Mol. Syndromol.* **2**, 186–201 (2012).
- Boeckers, T. M., Bockmann, J., Kreutz, M. R. & Gundelfinger, E. D. ProSAP/Shank proteins – a family of higher order organizing molecules of the postsynaptic density with an emerging role in human neurological disease. *J. Neurochem.* **81**, 903–910 (2002).
- Sheng, M. & Kim, E. The Shank family of scaffold proteins. *J. Cell Sci.* **113**, 1851–1856 (2000).
- Boccuto, L. *et al.* Prevalence of SHANK3 variants in patients with different subtypes of autism spectrum disorders. *Eur. J. Hum. Genet.* **21**, 310–316 (2013).
- Durand, C. M. *et al.* Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nature Genet.* **39**, 25–27 (2007).
- Gauthier, J. *et al.* Novel de novo SHANK3 mutation in autistic patients. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **150B**, 421–424 (2009).
- Moessner, R. *et al.* Contribution of SHANK3 mutations to autism spectrum disorder. *Am. J. Hum. Genet.* **81**, 1289–1297 (2007).
- Hamdan, F. F. *et al.* Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. *Am. J. Hum. Genet.* **88**, 306–316 (2011).
- Gauthier, J. *et al.* De novo mutations in the gene encoding the synaptic scaffolding protein SHANK3 in patients ascertained for schizophrenia. *Proc. Natl Acad. Sci. USA* **107**, 7863–7868 (2010).
- Wilson, H. L. *et al.* Molecular characterisation of the 22q13 deletion syndrome supports the role of haploinsufficiency of SHANK3/PROSAP2 in the major neurological symptoms. *J. Med. Genet.* **40**, 575–584 (2003).
- Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
- Yazawa, M. *et al.* Using induced pluripotent stem cells to investigate cardiac phenotypes in Timothy syndrome. *Nature* **471**, 230–234 (2011).
- Paşca, S. P. *et al.* Using iPSC-derived neurons to uncover cellular phenotypes associated with Timothy syndrome. *Nature Med.* **17**, 1657–1662 (2011).
- Yoo, A. S. *et al.* MicroRNA-mediated conversion of human fibroblasts to neurons. *Nature* **476**, 228–231 (2011).
- Dittgen, T. *et al.* Lentivirus-based genetic manipulations of cortical neurons and their optical and electrophysiological monitoring *in vivo*. *Proc. Natl Acad. Sci. USA* **101**, 18206–18211 (2004).
- Nathanson, J. L., Yanagawa, Y., Obata, K. & Callaway, E. M. Preferential labeling of inhibitory and excitatory cortical neurons by endogenous tropism of adeno-associated virus and lentivirus vectors. *Neuroscience* **161**, 441–450 (2009).
- Akhtar, M. W. *et al.* Histone deacetylases 1 and 2 form a developmental switch that controls excitatory synapse maturation and function. *J. Neurosci.* **29**, 8288–8297 (2009).
- Maunakea, A. K. *et al.* Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466**, 253–257 (2010).
- Rinaldi, T., Kulangara, K., Antonello, K. & Markram, H. Elevated NMDA receptor levels and enhanced postsynaptic long-term potentiation induced by prenatal exposure to valproic acid. *Proc. Natl Acad. Sci. USA* **104**, 13501–13506 (2007).
- Goold, C. P. & Nicoll, R. A. Single-cell optogenetic excitation drives homeostatic synaptic depression. *Neuron* **68**, 512–528 (2010).
- Marchetto, M. C. *et al.* A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. *Cell* **143**, 527–539 (2010).
- O'Kusky, J. R., Ye, P. & D'Ercole, A. J. Insulin-like growth factor-I promotes neurogenesis and synaptogenesis in the hippocampal dentate gyrus during postnatal development. *J. Neurosci.* **20**, 8435–8442 (2000).
- Tropea, D. *et al.* Partial reversal of Rett syndrome-like symptoms in MeCP2 mutant mice. *Proc. Natl Acad. Sci. USA* **106**, 2029–2034 (2009).
- Chen, D. Y. *et al.* A critical role for IGF-II in memory consolidation and enhancement. *Nature* **469**, 491–497 (2011).
- Bozdagi, O. *et al.* Haploinsufficiency of the autism-associated *Shank3* gene leads to deficits in synaptic function, social interaction, and social communication. *Mol. Autism* **1**, 15 (2010).
- Peça, J. *et al.* *Shank3* mutant mice display autistic-like behaviours and striatal dysfunction. *Nature* **472**, 437–442 (2011).
- Wang, X. *et al.* Synaptic dysfunction and abnormal behaviors in mice lacking major isoforms of *Shank3*. *Hum. Mol. Genet.* **20**, 3093–3108 (2011).
- Paoletti, P., Bellone, C. & Zhou, Q. NMDA receptor subunit diversity: impact on receptor properties, synaptic plasticity and disease. *Nature Rev. Neurosci.* **14**, 383–400 (2013).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We are grateful to participants and their families for their support; M. Adam for assistance with recruitment; to X. Jia, A. Cherry, C. Bangs, P. Jones, and J. Williams for assistance with tissue culture; P. Liao for help with multiplex ligation-dependent probe amplification (MLPA); M. Fabian for astrocyte preparations; H.N. Nguyen for consultations on the neural differentiation protocol and spectral karyotyping (SKY); V. Vu and G. Lin for help with data analysis; T. Sudhof, T. Boeckers, A. Grabruker, C. Garner and C. Sala for antibodies; R. Xavier for SHANK3 complementary DNA; R. Reijo-Pera and members of the Dolmetsch laboratory for commenting on the manuscript; E. Nigh for editing the manuscript. We also thank the Stanford Neuroscience Microscopy Service (supported by National Institutes of Health (NIH) NS069375). Support for this study came from the California Institute for Regenerative Medicine CIRM, the Autism Science Foundation and the Phelan-McDermid Syndrome Foundation (to A.S.), the Swiss National Science Foundation (to T.P.), the Japan Society for the Promotion of Research Abroad and American Heart Association (to M.Y.), the National Institute of Mental Health (NIMH) grant R33MH087898 (to J.F.H.); NIH Pioneer Award (5DP1OD3889), CIRM (grant RT2-01906) and Simons Foundation (to R.E.D.). We are also grateful for funding from the JDH research fund, N. Juaw, B. and F. Horowitz, M. McCaffrey, B. and J. Packard, P. Kwan and K. Wang, and the Flora foundation.

**Author Contributions** A.S. and R.E.D. designed experiments and wrote the manuscript; A.S. performed iPS cell maintenance, neural differentiation, electrophysiology, cloning and immunocytochemistry; O.S. maintained and characterized iPS cells, performed western blot and qRT-PCR; M.Y. generated and characterized iPS cells; T.P. performed multiplex single-cell qRT-PCR; V.S. performed teratoma assay; R.S. and A.K. performed qRT-PCR and data analysis; W.F., J.A.B. and J.F.H. recruited and characterized patients and performed the MLPA assay.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.E.D. ([Ricardo.dolmetsch@novartis.com](mailto:Ricardo.dolmetsch@novartis.com))

## METHODS

**Recruitment of patients with PMDS and skin biopsy procedure.** Participants diagnosed with 22q13 deletion syndrome were recruited from on-going research studies, the medical genetics service at Lucile Packard Children's Hospital, and interest groups for patients and families with genetic disorders associated with autistic spectrum manifestations. Consenting participants were evaluated with the Autism Diagnostic Observation Schedule (ADOS) and, when possible, the Stanford-Binet Intelligence Scales, Fifth Edition (SB5). In addition, parents of participants completed the Autism Diagnostic Interview-Revised (ADI-R) and questionnaires about the participant, including a medical history questionnaire and the Social Responsiveness Scale (SRS). All assessors were trained nationally and certified as reliable in the administration of ADOS and ADI-R assessments. Skin punch biopsies were performed as follows: after local anaesthesia with topical and injected xylocaine, a shallow, 3-mm punch biopsy sample was obtained from either the forearm or thigh.

This study was approved by both the Stanford University Institutional Review Board (IRB) and the Stem Cell Research Oversight Committee (SCRO).

**Generation and characterization of iPS cell lines.** Detailed procedures used for the generation, maintenance, and characterization of iPS cells were previously described<sup>12</sup>. Briefly, primary fibroblasts from patients with PMDS were acquired by skin-punch biopsy following approved SCRO and IRB protocols. Cells were propagated for 2–3 passages, plated at 100,000 cells per well in six-well plates, and infected with retroviruses carrying SOX2, OCT3/4, c-MYC and KLF4 transcription factors the next day. iPS cells colonies, normally detected 21–30 days after the first infection, were routinely maintained on irradiated CF1 feeder cells in iPS cell medium (DMEM/F12 (11330-032, Gibco), containing 20% Knockout SR (10828-028, Gibco), 1% MEM (11140, Gibco), 0.5% L-glutamine (25030, Gibco), 1% penicillin streptomycin (15140, Gibco), 0.1 mM  $\beta$ -mercaptoethanol (Sigma) and 10 ng ml<sup>-1</sup> basic fibroblast growth factor (R&D Systems)).

iPS cells were immunostained with antibodies against NANOG (AF1997, R&D Systems) and TRA-2-49/6E (developed by P. W. Andrews and obtained from the Development Studies Hybridoma Bank developed under the auspices of the National Institute of Child Health and Development (NICHD) and maintained in the University of Iowa, Department of Biological Sciences).

SKY analysis of all iPS cell lines was performed according to a standard procedure. Briefly, metaphases were extracted after overnight incubation in 0.03  $\mu$ g ml<sup>-1</sup> KaryoMAX Colcemid solution in iPS cell media. Chromosomes were isolated and stained with the SKYPaint probe mixture and CAD kit (Applied Spectral Imaging), following the manufacturer's protocol. Spreads were imaged on a Leica microscope (DM400B), equipped with SpectraCubeSD300 and a  $\times 63$  objective, using Case Data Manager 6.0 and HiSKY 6.0 software (Applied Spectral Imaging). Twenty spreads from each iPS cell line with well-separated chromosomes were imaged and analysed for visible chromosome abnormalities.

For the teratoma formation assay, iPS cells from a confluent 10-cm dish were enzymatically collected, washed and resuspended in 1.5 ml of PBS. Cells were then injected into the kidney capsules of SCID mice (Charles River Laboratories International). Visible tumours were dissected 3–8 weeks post-transplantation and fixed in 4% paraformaldehyde. Fixed samples were sent to the Cureline Biopathology Laboratory for paraffin embedding, sectioning and staining with haematoxylin and eosin. Sections were then examined for the presence of tissue representatives of all three germ layers.

Genomic DNA from fibroblasts and iPS cells was isolated using the DNeasy Blood & Tissue kit (Qiagen). Haploinsufficiency of *SHANK3* in fibroblasts and all iPS cells of patients, was verified using the SALSA MLPA kit (P188, MRC-Holland), following the manufacturer's protocol.

**Neural differentiation.** For neural differentiation, a combination of two protocols was used<sup>29,30</sup>. The rationale and detailed procedures are presented in Supplementary Fig. 3.

**qRT-PCR.** qRT-PCR was performed in a real-time thermal cycler (Mastercycler ep realplex, Eppendorf) using either FastStart Universal SYBR Green (Roche) or RT qPCR master mixes (SABiosciences). Total RNA was isolated using the RNeasy Mini kit (Qiagen). cDNA was synthesized with either the SuperScript III First-Strand Synthesis kit (Invitrogen) or the RT First Strand kit (SABiosciences). All primers are listed in Supplementary Table 5.

**Multiplex single-cell qRT-PCR.** Single cells were collected either by aspiration with a patch glass pipette filled with 4  $\mu$ l of sterile intracellular solution or by FACS sorting using a BD influx sorter (BD Biosciences) into 10  $\mu$ l of a pre-amplification mix containing 40 nM of all primers for the genes of interest and the following components of the CellsDirect One-Step qRT-PCR kit (Invitrogen): 2  $\times$  Reaction Mix, SuperScript III RT/Platinum Taq Mix. After collection, samples were reverse-transcribed and pre-amplified for 18 cycles. Pre-amplified samples were diluted (3  $\times$ ) with TE buffer and stored at  $-20^{\circ}\text{C}$ . Sample and assay (primer pairs) preparation for 96.96 Fluidigm Dynamic arrays was done according to the manufacturer's

recommendation. Briefly, each sample was mixed with 20  $\times$  DNA binding dye sample-loading reagent (Fluidigm), 20  $\times$  EvaGreen (Biotium) and TaqMan Gene Expression master mix (Applied Biosystems). Assays were mixed with 2  $\times$  assay loading reagent (Fluidigm) and TE to a final concentration of 5  $\mu$ M. The 96.96 Fluidigm Dynamic Arrays (Fluidigm) were primed and loaded on an IFC Controller HX (Fluidigm) and quantitative PCR experiments were run on a Biomark System for Genetic Analysis (Fluidigm). Melting curves were used to determine the specificity of each reaction. In addition to single-cell material, every experiment contained four standard dilutions of a mixed human cDNA library. Data were collected and analysed using Fluidigm Real-Time PCR Analysis software (v.2.1.3 and v.3.0.2). Threshold cycle (CT) as a measure for original template amount was used for quantification. To compare the data acquired from different chips, the  $C_t$  values were normalized to the median intensity of the entire chip. Because single-cell gene expression in mammals occurs in bursts and follows a binary distribution, we established 32 cycles (at least five standard deviations from the mean of the lowest expressed genes) as a general threshold for whether or not a gene is expressed.

**Electrophysiology.** All recordings were performed at room temperature ( $22\text{--}25^{\circ}\text{C}$ ) on GFP- or mKate2-expressing iPS cell-derived neurons on day 31–45 (separate cultures without astrocytes) or day 45–52 (co-cultured on rat cortical astrocytes). Briefly, neurons grown on cover slips or 35-mm plastic dishes were visualized with a  $\times 40$  air objective on an inverted Nikon microscope (Ellipse TE2000) and recorded using an EPC10 amplifier (HEKA). The following solutions were used (in mM): extracellular, 140 NaCl, 2.5 KCl, 2.5 CaCl<sub>2</sub>, 2 MgCl<sub>2</sub>, 1 NaH<sub>2</sub>PO<sub>4</sub>, 20 glucose, 10 HEPES, pH 7.4; intracellular, 135 CsMeS, 5 CsCl, 10 HEPES, 0.5 EGTA, 1 MgCl<sub>2</sub>, 4 Mg<sub>2</sub>ATP, 0.4 NaGTP, 5 QX-314, pH 7.4 with CsOH (for EPSCs, AMPA-, NMDA-, and  $\gamma$ -aminobutyric acid-current measurements) and 125 CsCl, 10 EGTA, 1 MgCl<sub>2</sub>, 0.1 CaCl<sub>2</sub>, 4 MgATP, 0.3 NaGTP, 10 HEPES, 5 QX-314 pH 7.2 with CsOH (for IPSCs measurements). Recording pipettes made of borosilicate glass (BF150-110-10, Sutter Instruments) had a resistance of 3–6 M $\Omega$  when filled with intracellular solution. The following combinations of antagonists were present in the extracellular solutions to isolate different postsynaptic currents (in  $\mu$ M): 20 bicuculline or 50 picrotoxin (spontaneous and evoked EPSCs), 10 NBQX and 50 2-amino-5-phosphonopentanoic acid (APV) (spontaneous and evoked IPSCs), 10 NBQX and 20 bicuculline or 50 picrotoxin (NMDA-current, focal application), 50 APV and 50 picrotoxin (AMPA-current, focal application), 10 NBQX and 50 APV ( $\gamma$ -aminobutyric acid-current, focal application). Spontaneous EPSCs and IPSCs were recorded for 3 min (1 min after break-in to block sodium current by QX-314) at holding potential  $-70$  mV. Recordings were filtered at 2 kHz and digitized at 4 kHz. Evoked postsynaptic currents were measured after spontaneous currents and induced by brief (1 ms) unipolar current pulses of various amplitudes (0.1–0.9 mA,  $\Delta = 0.05\text{--}0.1$  mA), delivered with a stimulating electrode (CBAEC75, FHC) positioned 100–150  $\mu$ m from the cell soma and connected to an external stimulator (A365, WPI). Recordings were filtered at 2 kHz and digitized at 10 kHz. NMDA-, AMPA- and  $\gamma$ -aminobutyric acid-currents were induced by a focal transient application of the following agonists (in  $\mu$ M): 100 NMDA and 10 Gly, 200 AMPA, or 100  $\gamma$ -aminobutyric acid. An injection pipette (patch pipette with resistance of 1–2 M $\Omega$  when filled with extracellular solution) was positioned about 20  $\mu$ m from the cell soma and connected to a pico-spritzer (PDES-02DX, npi Electronics) with 10 p.s.i. in-line pressure. In all experiments, access resistance was monitored throughout the experiment ( $\sim 15$  M $\Omega$ ), but not compensated. Cells were included for analysis only if the change of resistance was  $<25\%$  throughout the experiment.

Data were collected and initially analysed with Patchmaster software (HEKA). Further analysis was performed using Clampfit 10 (Molecular Devices), IgorPRO (Wavemetrics), Excel (Microsoft) and Prism (GraphPad). The baseline noise amplitude in our experiments was about 2.5 pA. Spontaneous EPSCs were detected using a template search procedure in Clampfit 10. A template for automatic EPSC detection was created by averaging 50 manually picked synaptic events. Spontaneous IPSCs were detected using MiniAnalysis software. The decay kinetics of NMDA-EPSCs was estimated using a weighted decay time constant ( $\tau_{\text{decay}}$ )<sup>31</sup>, calculated from the area under the normalized current on an interval from 16.4 ms ( $3 \times \tau_{\text{AMPA decay}}$ ) to 1.1 s after the peak of AMPA-EPSC.

**Western blot analysis.** Cells grown in 2-cm dishes were lysed in a standard RIPA buffer containing a protease inhibitor cocktail (Roche Diagnostics). Protein concentrations were measured using a BCA protein assay kit (Thermo). Proteins were denatured with Laemmli buffer at  $70^{\circ}\text{C}$  for 10 min, separated on a NuPAGE 4–12% Bis-Tris gel (Novex), and transferred to a polyvinylidene difluoride (PVDF) membrane. Membranes were blocked in 5% milk (1 h at room temperature) and blotted with the following primary antibodies (overnight at  $4^{\circ}\text{C}$ ): guinea pig anti-SHANK3 (1:700, gift from Carlo Sala), mouse anti-GLUN1 (1:1,000, SYSY), rabbit anti-GLUA1 (1:1,000, Upstate), mouse anti-GLUA2 (1:1,000, NeuroMab), and mouse anti-GAPDH (MS, 1:1,000, Ambion). To control loading protein amounts, we first quantified the relative concentrations of GAPDH in all lysates. Afterwards,

15–25 µg of proteins were loaded into a gel and intensities of appropriate bands were measured and normalized to the intensity of the GAPDH band in the same lane. Bands were visualized and quantified using the FluorChemQ System (Alpha Innotech).

**Immunocytochemistry, imaging and synaptic assays.** Cells were fixed in 4% paraformaldehyde (15 min at room temperature), washed (3×) with PBS glycine, permeabilized with 0.5% Nonidet P40 (7 min at room temperature), washed (3×) with PBS, and blocked in 3% BSA (overnight at 4 °C). Primary antibodies were diluted in 3% BSA and applied overnight at 4 °C. The following primary antibodies were used: chicken anti-GFP (Abcam, 1:1,000), rabbit anti-tRFP (Axxora, 1:1,000), guinea pig anti-SHANK3 (produced by TM Boeckers, 1:800), mouse anti-SYN1 (SYSY, 1:500), guinea pig anti-HOMER1 (SYSY, 1:200), mouse anti-PSD95 (Abcam, 1:200), rat anti-CTIP2 (Abcam, 1:200), mouse anti-SATB2 (Abcam, 1:10), mouse anti-GAD67, and anti-nuclei (Millipore, 1:100). Secondary antibodies were diluted (1:500–1,000) in 3% BSA and applied for 1 h at room temperature. The following secondary antibodies were used: Alexa-405, -488, -594 and -647 (Invitrogen). Neurons were visualized using either a Zeiss spinning disc confocal microscope, equipped with 405/488/594/640 lasers (Perkin Elmer) or a laser-scanning confocal microscope equipped with 405/488/594/633 lasers (Stanford Neuroscience Microscopy Services), using either Volocity (Improvision) or Zenn (Zeiss) software. Z-stack images of fluorescent cells were taken with a 0.2–0.7 µm interval using ×100 or ×63 objectives that cover an 80 × 80 or 150 × 150 µm<sup>2</sup> area, correspondingly. Images within an experiment were acquired with identical acquisition settings. Approximately 20 images (10 GFP<sup>+</sup> and 10 mKate2<sup>+</sup> neurons) were collected from each coverslip. Visualization and analysis of all images were performed using Volocity 5.1 and ImageJ. For quantification of total SHANK3 expression, a three-dimensional mask of a neuron was created based on GFP or mKate2 fluorescence and the mean SHANK3 intensity within the masked area was calculated. For quantification of the number of synapses, Z-stacks were merged using maximum intensity readings for each of the channels and a two-dimensional mask of a neuron was created based on the GFP or mKate2 fluorescence reading. Presynaptic (SYN1) and postsynaptic (SHANK3, HOMER1 or PSD-95) puncta within a mask were found either manually using a cell counter plug-in for ImageJ or automatically using a custom measurement protocol created in Volocity. During automatic quantification, once settings for puncta detection were chosen, they were uniformly applied for all images within an experiment. Results of automatic puncta quantification were always validated by visual inspection and manual counting for a subset of the images. All cover slips were coded and processed in a blinded manner. To combine data acquired from different coverslips in different experiments, we normalized all puncta counts to the median number of puncta detected on control neurons for a particular cover slip.

**Construction of SHANK3 lentiviral vector.** The translated region of the rat SHANK3 cDNA (NM\_021676.1, generously provided by R. Xavier) was isolated in pieces by PCR and flanked with NdeI (5') and ClaI (3') restriction enzyme sites, and then subcloned into the pSC-B-amp/kan vector using the StrataClone Blunt PCR Cloning kit (Stratagene). Similarly, a DNA fragment of EGFP flanked with

NotI (3' UTR), NdeI (5' TR before stop codon), and ClaI (3' UTR) restriction enzyme sites was obtained by PCR and subcloned into a pSC-B-amp/kan vector. Afterwards, NotI-EGFP-ClaI (0.7 kb) and NdeI-SHANK3-ClaI (5.2 kb) fragments were isolated, gel-purified, and sequentially inserted into the NotI and ClaI sites and then into an NdeI- and ClaI-digested pHAGE-EF1a lentiviral vector backbone<sup>32</sup> to generate the pHAGE-EF1a-EGFP-SHANK3 (13 kb) vector. The sequence was confirmed by sequencing.

**Lentivirus production, titration and infection.** Lentiviruses were generated with a five-plasmid transfection system using Eugene 6 transfection reagent (Roche) in 293T cells as previously described<sup>32</sup>. 48 h after transfection, supernatants were collected at 12-h intervals, 4–5 times in total. Viral particles were concentrated using Lenti-X concentrator (Clontech), aliquoted (200 µl), and stored at –80 °C.

Different amounts of virus (5–100 µl) were placed on ~80% confluent HEK293T cells for 3–4 days. Approximately 80% of the cells in the 35-mm plate were fluorescent upon addition of 100 µl of virus. We used 100 µl of GFP-SHANK3 virus to transduce human iPS cell-derived neurons (Day 25–30) growing in 35-mm culture plates. The medium was replaced 10–20 h later. Neurons were transferred onto rat primary astrocytes after 5–10 days.

**Drugs and treatment procedure.** The following agents were tested: 0.1 µM trichostatin A (TSA, Tocris), 1 mM valproic acid (VPA, Calbiochem), 3 µM nifedipine (Nif, Tocris), 20 ng ml<sup>–1</sup> insulin growth factor 1 (IGF1, R&D Systems), and 20 ng ml<sup>–1</sup> insulin growth factor 2 (IGF2, R&D Systems). Concentrations were chosen on the basis of previous reports of their use available in the literature. Effects of TSA, VPA, and Nif were assessed after 24–36 h of treatment and IGF1 and IGF2 after 2–3 weeks, with approximately 25% of the media changed every other day.

**Data collection and statistics.** All experiments were replicated at least three times using iPS cell lines indicated in Supplementary Table 3. The sample size and description of the sample collection are reported in each figure legend. The number of samples was selected based on the variability of the particular assay. Samples were excluded from analysis only when they were clear outliers (identified visually and confirmed with the Grubbs' test) or were erroneously generated as a result of a technical problem. Wells with neurons were randomly selected for an assay/drug treatment, coded, and processed in a blinded manner. Statistical tests used for comparison are reported in each figure legends. For data sets where parametric tests were used for analysis, the normal distribution of data was confirmed using the Kolmogorov–Smirnov test.

29. Chambers, S. M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nature Biotechnol.* **27**, 275–280 (2009).
30. Gaspard, N. *et al.* Generation of cortical neurons from mouse embryonic stem cells. *Nature Protocols* **4**, 1454–1463 (2009).
31. Bellone, C. & Nicoll, R. A. Rapid bidirectional switching of synaptic NMDA receptors. *Neuron* **55**, 779–785 (2007).
32. Sommer, C. A. *et al.* Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. *Stem Cells* **27**, 543–549 (2009).



# Paneth cells as a site of origin for intestinal inflammation

Timon E. Adolph<sup>1\*</sup>, Michal F. Tomczak<sup>2\*</sup>, Lukas Niederreiter<sup>1\*</sup>, Hyun-Jeong Ko<sup>2†\*</sup>, Janne Böck<sup>3</sup>, Eduardo Martinez-Naves<sup>4</sup>, Jonathan N. Glickman<sup>5</sup>, Markus Tschurtschenthaler<sup>1,6</sup>, John Hartwig<sup>7</sup>, Shuhei Hosomi<sup>2</sup>, Magdalena B. Flak<sup>2</sup>, Jennifer L. Cusick<sup>2</sup>, Kenji Kohno<sup>8</sup>, Takao Iwawaki<sup>9,10</sup>, Susanne Billmann-Born<sup>3</sup>, Tim Raine<sup>1</sup>, Richa Bharti<sup>3</sup>, Ralph Lucius<sup>11</sup>, Mi-Na Kweon<sup>12</sup>, Stefan J. Marciniak<sup>13</sup>, Augustine Choi<sup>14</sup>, Susan J. Hagen<sup>15</sup>, Stefan Schreiber<sup>3</sup>, Philip Rosenstiel<sup>3</sup>, Arthur Kaser<sup>1\*</sup> & Richard S. Blumberg<sup>2\*</sup>

**The recognition of autophagy related 16-like 1 (ATG16L1) as a genetic risk factor has exposed the critical role of autophagy in Crohn's disease<sup>1</sup>. Homozygosity for the highly prevalent ATG16L1 risk allele, or murine hypomorphic (HM) activity, causes Paneth cell dysfunction<sup>2,3</sup>. As *Atg16l1*<sup>HM</sup> mice do not develop spontaneous intestinal inflammation, the mechanism(s) by which ATG16L1 contributes to disease remains obscure. Deletion of the unfolded protein response (UPR) transcription factor X-box binding protein-1 (*Xbp1*) in intestinal epithelial cells, the human orthologue of which harbours rare inflammatory bowel disease risk variants, results in endoplasmic reticulum (ER) stress, Paneth cell impairment and spontaneous enteritis<sup>4</sup>. Unresolved ER stress is a common feature of inflammatory bowel disease epithelium<sup>4,5</sup>, and several genetic risk factors of Crohn's disease affect Paneth cells<sup>2,4,6–9</sup>. Here we show that impairment in either UPR (*Xbp1*<sup>ΔIEC</sup>) or autophagy function (*Atg16l1*<sup>ΔIEC</sup> or *Atg7*<sup>ΔIEC</sup>) in intestinal epithelial cells results in each other's compensatory engagement, and severe spontaneous Crohn's-disease-like transmural ileitis if both mechanisms are compromised. *Xbp1*<sup>ΔIEC</sup> mice show autophagosome formation in hypomorphic Paneth cells, which is linked to ER stress via protein kinase RNA-like endoplasmic reticulum kinase (PERK), elongation initiation factor 2α (eIF2α) and activating transcription factor 4 (ATF4). Ileitis is dependent on commensal microbiota and derives from increased intestinal epithelial cell death, inositol requiring enzyme 1α (IRE1α)-regulated NF-κB activation and tumour-necrosis factor signalling, which are synergistically increased when autophagy is deficient. ATG16L1 restrains IRE1α activity, and augmentation of autophagy in intestinal epithelial cells ameliorates ER stress-induced intestinal inflammation and eases NF-κB overactivation and intestinal epithelial cell death. ER stress, autophagy induction and spontaneous ileitis emerge from Paneth-cell-specific deletion of *Xbp1*. Genetically and environmentally controlled UPR function within Paneth cells may therefore set the threshold for the development of intestinal inflammation upon hypomorphic ATG16L1 function and implicate ileal Crohn's disease as a specific disorder of Paneth cells.**

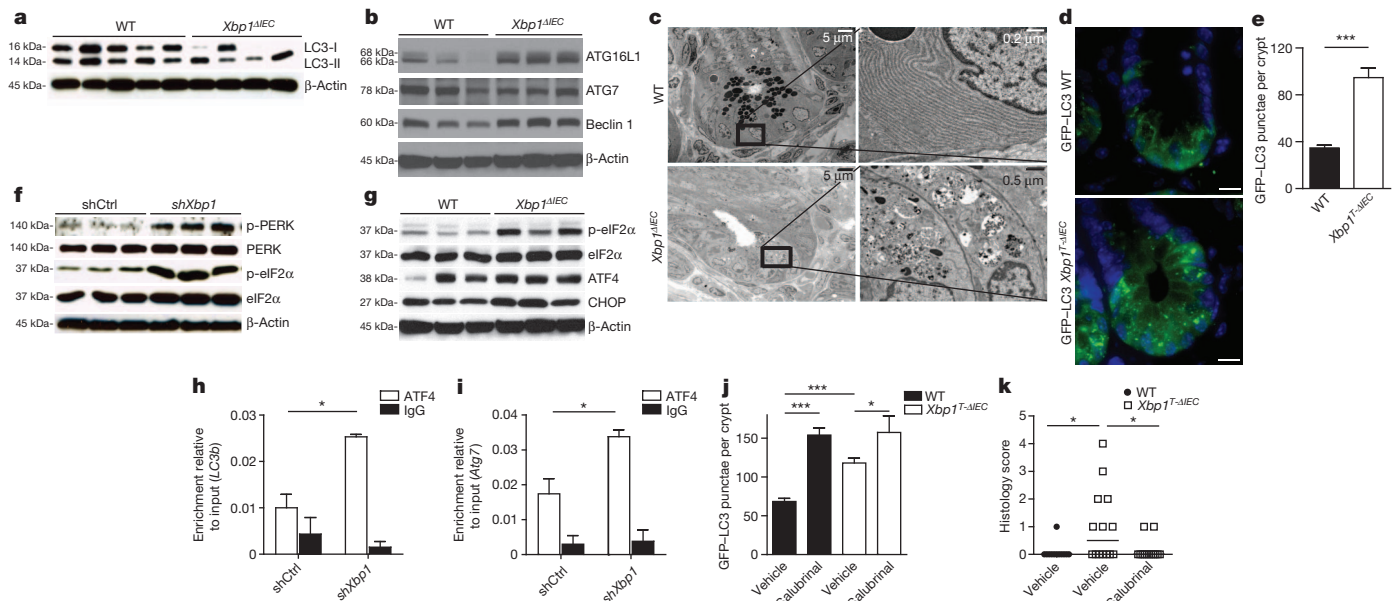
The UPR and autophagy are integrally linked pathways<sup>10</sup>. To investigate their relationship in the intestinal epithelium, we stably transduced the small intestinal epithelial cell line MODE-K with a short hairpin *Xbp1* (*shXbp1*) lentiviral vector (Extended Data Fig. 1a)<sup>4</sup>. We observed increased levels of ATF7, beclin 1 (Extended Data Fig. 1b) and phosphatidylethanolamine (PE)-conjugated microtubule-associated protein 1 light chain 3-β (LC3-II) relative to LC3-I compared to control-silenced (*shCtrl*) cells (Extended Data Fig. 1c). Increased autophagic flux accounted

for this, given increased levels of LC3-II relative to LC3-I observed after inhibition of autophagosome-lysosome fusion by bafilomycin<sup>11</sup> (Extended Data Fig. 1d). Increased GFP-LC3 punctae were seen in *shXbp1* compared to *shCtrl* MODE-K cells transfected with a green fluorescent protein (GFP)-LC3-expressing vector<sup>12</sup> (Extended Data Fig. 1e, f), as well as increased numbers of autophagic vacuoles in *shXbp1* compared to *shCtrl* cells (Extended Data Fig. 1g, h). Accordingly, isolated primary intestinal epithelial cells from the small intestine of villin (V)-*cre*<sup>+</sup>; *Xbp1*<sup>fl/fl</sup> (hereafter called *Xbp1*<sup>ΔIEC</sup>)<sup>4</sup> mice backcrossed onto C57BL/6 (B6) background exhibited nearly complete consumption of LC3-I and a relative increase in LC3-II (Fig. 1a and Extended Data Fig. 1i), stable amounts of ATF7 (presumably reflecting a combination of increased production and consumption), elevated levels of ATG16L1 and beclin 1 (Fig. 1b and Extended Data Fig. 1j), and autophagosomes and degradative autophagic vacuoles consistent with autophagy induction in hypomorphic<sup>4</sup> Paneth cells, and to a lesser extent goblet cells (data not shown), compared to *V-cre*<sup>+</sup>; *Xbp1*<sup>fl/fl</sup> (hereafter called wild type) mice (Fig. 1c and Extended Data Fig. 1k). To gain temporal control of *Xbp1* deletion and the ability to monitor autophagy *in situ* and exclude the role of chronic inflammation<sup>4</sup> in this induction, we generated *V-creER*<sup>+</sup>; *Xbp1*<sup>fl/fl</sup> (hereafter called *Xbp1*<sup>T-ΔIEC</sup>) mice on a B6 background crossed to *GFP-LC3* transgenic mice<sup>12</sup>. Three days after tamoxifen-induced *Xbp1* deletion (Extended Data Fig. 2a), although mature Paneth cells remained present with little detectable inflammation (data not shown), punctate GFP signal accumulation was greatest at the bottom of the crypts of Lieberkühn (Fig. 1d, e), and co-localized with lysozyme-positive Paneth cells (Extended Data Fig. 2b). Purified crypts of *Xbp1*<sup>T-ΔIEC</sup> mice revealed increased LC3-I/II conversion and reduced p62 compared to wild-type mice (Extended Data Fig. 2c). Thus, *Xbp1* loss in intestinal epithelial cells induced autophagy most notably in Paneth cells.

Although *Xbp1*-deficient intestinal epithelial cells exhibited broad evidence of ER stress<sup>4</sup> (Extended Data Fig. 2d–j), an examination of *shXbp1*, relative to *shCtrl*, MODE-K cells demonstrated a particularly significant increase in phosphorylated (p)-PERK and its substrate p-eIF2α (Fig. 1f), with the latter reversed by *Perk* (also called *Eif2ak3*) silencing, identifying it as the factor responsible for p-eIF2α formation (Extended Data Fig. 3a). Increased p-eIF2α was also detected in primary intestinal epithelial cells of *Xbp1*<sup>ΔIEC</sup> (Fig. 1g and Extended Data Fig. 3b) and *Xbp1*<sup>T-ΔIEC</sup> mice (Extended Data Fig. 3c). Consistent with the involvement of PERK-eIF2α in autophagy induction, ATF4, a transcriptional effector of this pathway, and its transcriptional target, C/EBP-homologous protein (CHOP; encoded by *Ddit3*), were increased in primary intestinal

<sup>1</sup>Division of Gastroenterology and Hepatology, Department of Medicine, Addenbrooke's Hospital, University of Cambridge, Cambridge CB2 0QQ, UK. <sup>2</sup>Division of Gastroenterology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, 75 Francis Street, Boston, Massachusetts 02115, USA. <sup>3</sup>Institute for Clinical Molecular Biology, Christian-Albrechts-Universität zu Kiel, D-24105 Kiel, Germany. <sup>4</sup>Department of Microbiology and Immunology, Facultad de Medicina, Universidad Complutense de Madrid, 28040 Madrid, Spain. <sup>5</sup>GI Pathology Division, Miraca Life Sciences, Newton, Massachusetts 02464, USA. <sup>6</sup>Department of Medicine, Innsbruck Medical University, A-6020 Innsbruck, Austria. <sup>7</sup>Translational Medicine Division, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, 75 Francis Street, Boston, Massachusetts 02115, USA. <sup>8</sup>Laboratory of Molecular and Cell Genetics, Graduate School of Biological Sciences, Nara Institute of Science and Technology (NAIST), 8916-5 Takayama, Ikoma, Nara 630-0192, Japan. <sup>9</sup>Advanced Scientific Research Leaders Development Unit, Gunma University 3-39-22 Showa-machi, Maebashi, Gunma 371-8511, Japan. <sup>10</sup>Iwawaki Initiative Research Unit, Advanced Science Institute, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan. <sup>11</sup>Anatomical Institute, Christian-Albrechts-Universität zu Kiel, D-24098 Kiel, Germany. <sup>12</sup>Mucosal Immunology Section, Laboratory Science Division, International Vaccine Institute, Seoul 151-818, South Korea. <sup>13</sup>Department of Medicine, University of Cambridge, Cambridge Institute for Medical Research (CIMR), Wellcome Trust/MRC Building, Hills Road, Cambridge CB2 0XY, UK. <sup>14</sup>Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, 75 Francis Street, Boston, Massachusetts 02115, USA. <sup>15</sup>Department of Surgery, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02215, USA. <sup>†</sup>Present address: Laboratory of Microbiology and Immunology, College of Pharmacy, Kangwon National University, Chuncheon 200-701, South Korea.

\*These authors contributed equally to this work.

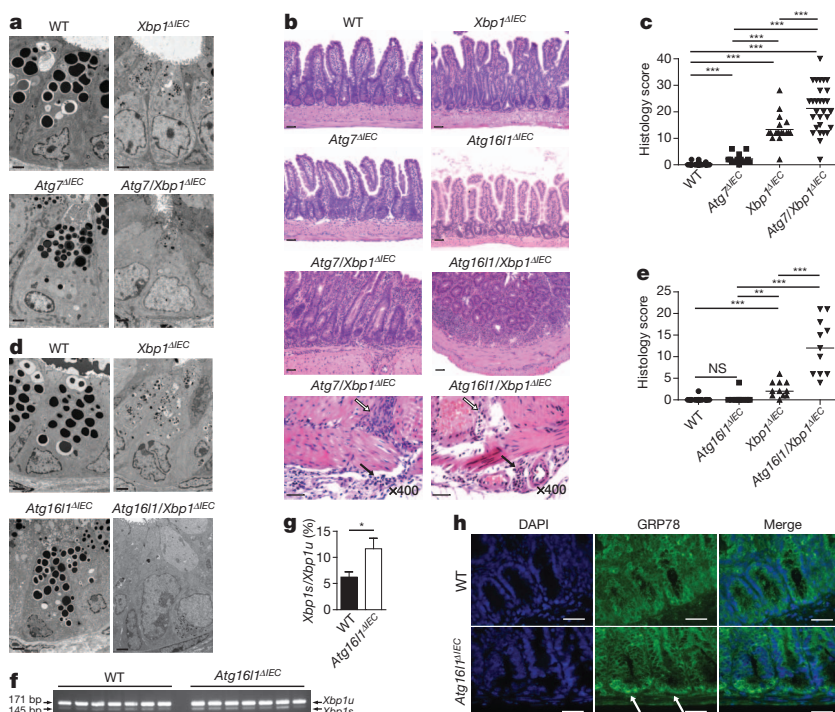


**Figure 1 | PERK-eIF2 $\alpha$  signalling induces autophagy in *Xbp1*-deficient intestinal epithelial cells.** **a, b**, Immunoblot for LC3 conversion in isolated primary intestinal epithelial cells (**a**) ( $n = 5/4$ ) and for autophagy proteins in primary intestinal epithelial cell scrapings (**b**) ( $n = 3$ ). WT, wild type. **c**, Transmission electron microscopy (TEM) of crypts. Note autophagic vacuoles in various stages of evolution in *Xbp1<sup>ΔIEC</sup>* hypomorphic Paneth cells. **d, e**, Crypt showing GFP-LC3 punctae (**d**), quantified in **e** ( $n = 10$ ; unpaired Student's *t*-test; mean  $\pm$  s.e.m.). Scale bar, 5  $\mu$ m. **f, g**, Immunoblot of silenced MODE-K cells (**f**) and primary intestinal epithelial cell scrapings (**g**) for the

epithelial cells of *Xbp1<sup>ΔIEC</sup>* mice (Fig. 1g and Extended Data Fig. 3b), and chromatin-immunoprecipitation (ChIP) with anti-ATF4 demonstrated increased binding to the *Map1lc3b* (*LC3b*) (Fig. 1h) and *Atg7* (Fig. 1i) promoters, both of which contain ATF4 binding sites<sup>13</sup>, in *shXbp1* relative to shCtrl MODE-K cells. ATG7 is essential for the formation of the ATG12-ATG5 conjugate during autophagy<sup>10,14</sup>. *shXbp1* MODE-K cells showed increased *LC3b* and *Atg7* expression compared to shCtrl MODE-K cells (Extended Data Fig. 3d), and *Perk* co-silencing abrogated ATG7 induction observed in *shXbp1* compared to shCtrl MODE-K cells

(Extended Data Fig. 3a). Salubrinal, a selective inhibitor of eIF2 $\alpha$  dephosphorylation<sup>15</sup> (Extended Data Fig. 2a), increased the accumulation of GFP-LC3 punctae primarily in Paneth cells, in both *Xbp1*-sufficient and -deficient intestinal epithelial cells (Fig. 1j and Extended Data Fig. 2b and 3e), and provoked increased levels of LC3-II relative to LC3-I (Extended Data Fig. 3f) and CHOP (Extended Data Fig. 3f) in *Xbp1<sup>ΔIEC</sup>* mice, along with, importantly, an amelioration of the acute enteritis (Fig. 1k and Extended Data Fig. 3g). Similarly, silencing of growth arrest and DNA-damage-inducible protein 34 (*Gadd34*; also

PERK-eIF2 $\alpha$  branch ( $n = 3$ ). **h, i**, Promoter sequence qPCR for *Map1lc3b* (*LC3b*) (**h**) and *Atg7* (**i**) after anti-ATF4 ChIP (unpaired Student's *t*-test; mean  $\pm$  s.e.m.). **j**, GFP-LC3 punctae per crypt after treatment with tamoxifen, and vehicle or salubrinal ( $n = 10$ ; one-way ANOVA with post-hoc Bonferroni; mean  $\pm$  s.e.m.). **k**, Enteritis histology score after salubrinal and tamoxifen co-treatment ( $n = 12/14/13$ ; median shown; Kruskal-Wallis with post-hoc Holm's-corrected Mann-Whitney *U*-test). Results represent three (**a, f, g**) or two (**c, e, h, i**) independent experiments. \* $P < 0.05$ , \*\*\* $P < 0.001$ .



**Figure 2 | Impairment of ER stress-induced compensatory autophagy results in severe transmurial inflammation.** **a**, TEM of crypts of Lieberkühn ( $n = 2$ ). Scale bars, 2  $\mu$ m. **b**, Representative haematoxylin and eosin stainings. Note transmurial inflammation extending through muscularis propria (white arrow) into serosa (black arrow) in *Atg7/Xbp1<sup>ΔIEC</sup>* and *Atg16l1/Xbp1<sup>ΔIEC</sup>* mice scored in **c** and **e**. Scale bars, 50  $\mu$ m, except for lower two panels (10  $\mu$ m). **c**, Enteritis histology score ( $n = 26/12/18/27$ ; 10–18 weeks; median shown; Kruskal-Wallis with post-hoc Holm's-corrected Mann-Whitney *U*-test). **d**, Crypt TEM in indicated genotypes ( $n = 2$ ). Scale bars, 2  $\mu$ m. **e**, Enteritis histology score ( $n = 11$ ; 18 weeks; median shown; Kruskal-Wallis with post-hoc Holm's-corrected Mann-Whitney *U*-test). NS, not significant. **f**, *Xbp1* mRNA splicing (*Xbp1u*, unspliced; *Xbp1s*, spliced) of crypts; densitometry in **g** ( $n = 7$ ; unpaired Student's *t*-test; mean  $\pm$  s.e.m.). **h**, GRP78 (green) immunofluorescence, white arrows indicate GRP78<sup>+</sup> crypts (DAPI, blue;  $n = 5$ ). Scale bars, 10  $\mu$ m. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .



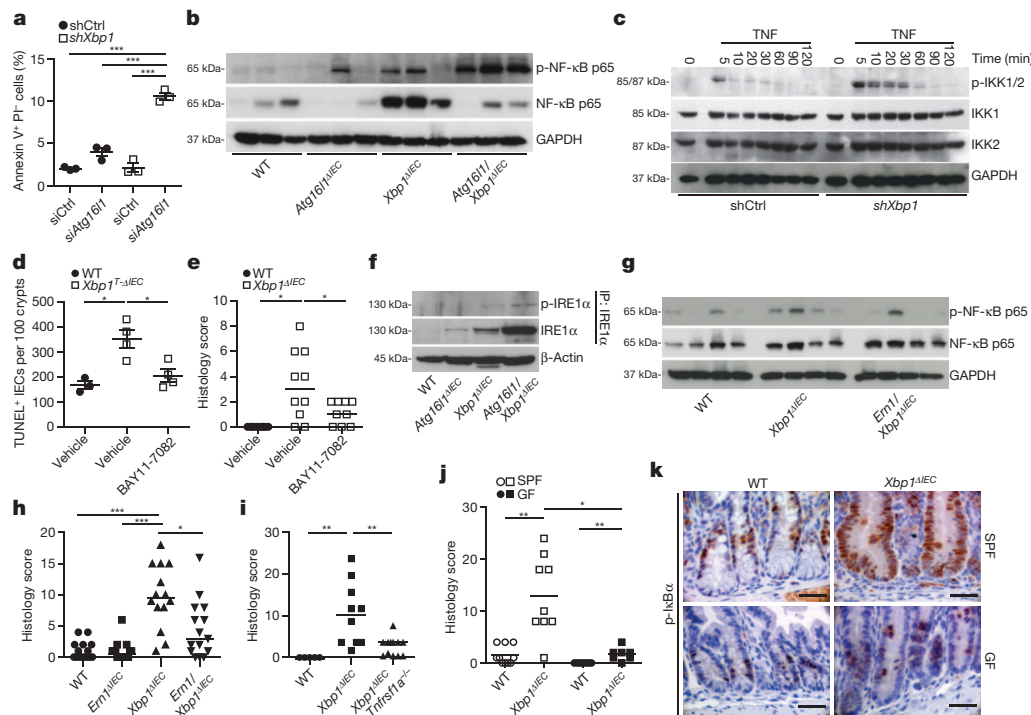
called *Ppp1r15a*), part of the protein phosphatase 1 complex that dephosphorylates eIF2 $\alpha$ <sup>16</sup>, led to increased eIF2 $\alpha$  phosphorylation and ATG7 expression in *shXbp1* compared to *shCtrl* MODE-K cells (Extended Data Fig. 3h, i). *Xbp1<sup>ΔIEC</sup>;Gadd34<sup>+/-</sup>* mice with hypomorphic GADD34 function exhibited increased p-eIF2 $\alpha$  and ATG7 in purified crypt epithelial cells compared to *Xbp1<sup>ΔIEC</sup>;Gadd34<sup>+/+</sup>* mice (Extended Data Fig. 3j). Thus, PERK–p-eIF2 $\alpha$  is a critical mediator of UPR-induced autophagy primarily in Paneth cells consequent to XBP1 deficiency.

These studies led us to propose that autophagy may function as a compensatory mechanism in intestinal epithelial cells upon sustained ER stress. We therefore generated *V-cre<sup>+</sup>;Atg7<sup>fl/fl</sup>;Xbp1<sup>fl/fl</sup>* (*Atg7/Xbp1<sup>ΔIEC</sup>*) mice<sup>14</sup>. Intestinal epithelial cells from *Atg7<sup>ΔIEC</sup>* mice lacked LC3-II formation and the ATG12–ATG5–ATG16L1 complex (Extended Data Fig. 4a). *Atg7/Xbp1<sup>ΔIEC</sup>* mice demonstrated a complete absence of UPR-induced autophagy (Fig. 2a and Extended Data Fig. 4a–c), and a remarkable worsening of ileitis compared to *Xbp1<sup>ΔIEC</sup>* mice. In notable contrast to *Xbp1<sup>ΔIEC</sup>* mice, where inflammation was limited to the mucosa, >70% of *Atg7/Xbp1<sup>ΔIEC</sup>* mice developed discontinuous submucosal or transmural inflammation, characterized by acute and chronic inflammation extending in an abrupt knife-like fashion to muscularis propria and serosa, closely resembling the early fissuring ulcerations and fistulous tracts observed in human Crohn's disease (Fig. 2b, c and Extended Data Fig. 4d). In contrast to *Xbp1<sup>ΔIEC</sup>* mice, enteritis in *Atg7/Xbp1<sup>ΔIEC</sup>* mice progressed over the 18-week observation period such that at this time point all animals exhibited submucosal or transmural disease (Extended Data Fig. 4e, f).

ATG16L1 is a major genetic risk factor for Crohn's disease<sup>1,17</sup>, especially ileal Crohn's disease<sup>18</sup>. Complex formation of ATG16L1 protein with ATG12–ATG5 defines the site of LC3 PE conjugation during autophagosome formation<sup>19,20</sup>. ATG16L1 protein expression was markedly

increased in *Xbp1<sup>ΔIEC</sup>* compared to wild-type primary intestinal epithelial cells (Fig. 1b and Extended Data Fig. 1j), presumably consequent to PERK/eIF2 $\alpha$ /ATF4-dependent transactivation of *Atg7* and *LC3b* promoters and stabilization by the ATG7-induced ATG12–ATG5 complex<sup>21</sup>. We therefore developed mice with a floxed *Atg16l1* allele that would allow for intestinal-epithelial-cell-specific deletion via *V-cre* (*Atg16l1<sup>ΔIEC</sup>*; Extended Data Fig. 4g–i). Paneth cells in *Atg16l1<sup>ΔIEC</sup>* mice demonstrated a reduction in their overall size and number of granules, similar to gene-trap-targeted *Atg16l1<sup>HM</sup>* mice<sup>2,3</sup> (Extended Data Fig. 4j–n). Intestinal epithelial cells from *Atg16l1<sup>ΔIEC</sup>* mice, compared to wild-type mice, exhibited reduced expression of ATG7 and the ATG12–ATG5 conjugate (Extended Data Fig. 5a), along with disruption of the secretory pathway with a distended ER, reduced size and number of secretory granules, a loss of homeostatic autophagy (Fig. 2d and Extended Data Fig. 5b, c) and increased p62 immunoreactivity in crypts (Extended Data Fig. 5d). To address the role of ATG16L1 under ER stress conditions, we generated *V-cre<sup>+</sup>;Atg16l1<sup>fl/fl</sup>;Xbp1<sup>fl/fl</sup>* (*Atg16l1/Xbp1<sup>ΔIEC</sup>*) mice. *Atg16l1/Xbp1<sup>ΔIEC</sup>* mice, which lacked UPR-induced autophagy (Extended Data Fig. 5b, c), developed severe spontaneous ileitis compared to *Xbp1<sup>ΔIEC</sup>* or *Atg16l1<sup>ΔIEC</sup>* mice, with discontinuous submucosal or transmural inflammation in >70% of 18-week-old animals (Fig. 2b, e and Extended Data Fig. 5e, f) with features similar to those observed in *Atg7/Xbp1<sup>ΔIEC</sup>* mice, and present in two distinct animal facilities (Fig. 2b, e and Extended Data Fig. 5g). This phenotype highlights the important compensatory role played by autophagy and in particular ATG16L1 in defending against inflammation arising from unabated ER stress precisely in the small intestinal epithelium as a consequence of XBP1 deficiency.

ATG7 hypofunction in hepatocytes can induce ER stress<sup>22</sup>, raising the possibility that cross-talk between the UPR and autophagy may be



**Figure 3 | Autophagy restrains IRE1 $\alpha$ -mediated NF- $\kappa$ B activation in *Xbp1*-deficient epithelium.** **a**, *shCtrl* or *shXbp1* MODE-K cells were co-silenced for *Atg16l1* (*siAtg16l1*) or with scrambled siRNA (*siCtrl*), and analysed by flow cytometry for annexin V and propidium iodide (PI;  $n = 3$ ; one-way ANOVA with post-hoc Bonferroni; mean  $\pm$  s.e.m.). **b**, Immunoblot of primary intestinal epithelial cell scrapings ( $n = 3$ ). **c**, Immunoblot of cytoplasmic extracts from *shCtrl* or *shXbp1* MODE-K cells after TNF stimulation. **d**, TUNEL<sup>+</sup> intestinal epithelial cells (IECs) per 100 crypts after BAY11-7082 or vehicle treatment ( $n = 3/4/4$ ; one-way ANOVA with post-hoc Holm's-corrected unpaired Student's *t*-test; mean  $\pm$  s.e.m.). **e**, Enteritis histology score of mice treated with BAY11-7082 or vehicle ( $n = 10/10/9$ ; median shown; Kruskal–Wallis with post-hoc Holm's-corrected

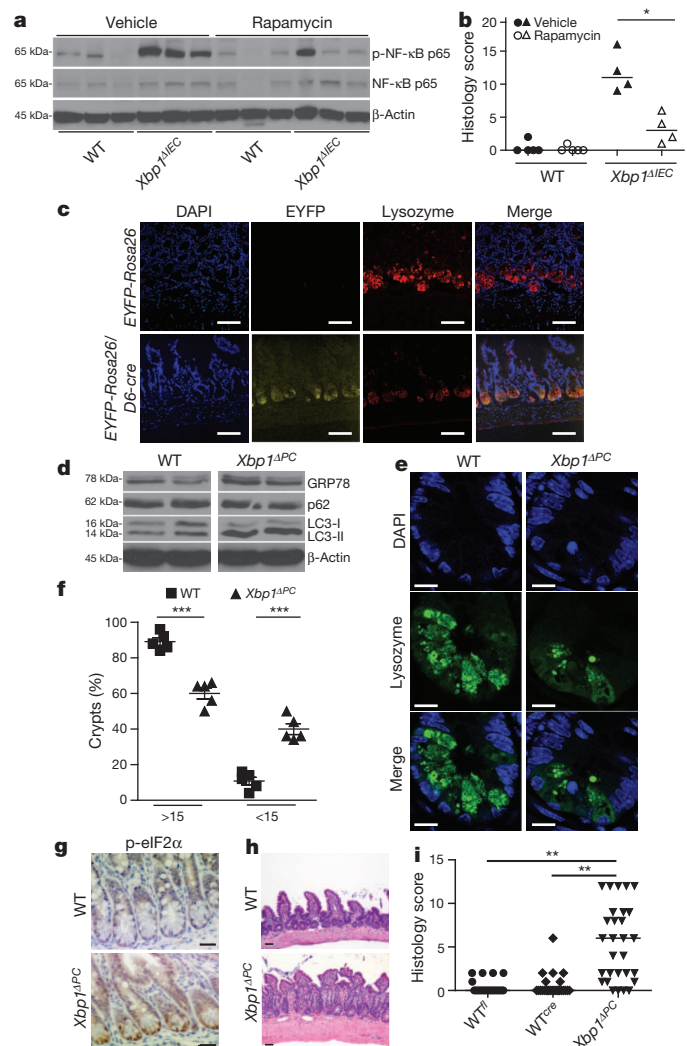
Mann–Whitney *U*-test). **f**, Immunoblot of intestinal epithelial cell scrapings for (p-)IRE1 $\alpha$  after IRE1 $\alpha$  immunoprecipitation (IP).  $\beta$ -Actin, loading control of whole lysates. **g**, Immunoblot of primary intestinal epithelial cell scrapings ( $n = 4$ ). **h**, **i**, Enteritis histology score of indicated genotypes (**h**,  $n = 15/16/14/15$ ; **i**,  $n = 5/10/12$ ; median shown; Kruskal–Wallis with post-hoc Holm's-corrected Mann–Whitney *U*-test). **j**, Enteritis histology score of specific pathogen-free (SPF) and germ free (GF) housed mice ( $n = 10/9/7/7$ ; median shown; Kruskal–Wallis with post-hoc Holm's-corrected Mann–Whitney *U*-test). **k**, Representative images of p-IkB $\alpha$  immunoreactivity under conditions as in **j** ( $n = 4$ ). Scale bars, 20  $\mu$ m. Results represent four (**f**), three (**c**) or two (**a**, **b**) independent experiments. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .



bi-directional in intestinal epithelial cells. Isolated *Atg16l1<sup>AIEC</sup>* crypts exhibited increased *Xbp1* splicing compared to wild type (Fig. 2f, g) and increased *Grp78* (also called *Hspa5*) expression (Extended Data Fig. 5h) localized to the crypt bottom (Fig. 2h). Dextran sodium sulphate, a colitis model involving ER stress in intestinal epithelial cells that can be treated with ER stress-relieving chaperones<sup>23,24</sup>, induced more inflammation in *Atg16l1<sup>AIEC</sup>* compared to wild-type mice (Extended Data Fig. 5i–l), similar to *Atg16l1<sup>HM</sup>* mice<sup>3</sup>. Thus, disturbances in autophagy within intestinal epithelial cells also affect the UPR, and autophagy-associated factors such as ATG16L1 endow intestinal epithelial cells with the ability to mitigate ER stressors that are commonplace at the mucosal surface<sup>25</sup>.

We next turned our attention to mechanisms by which autophagy counteracts ER stress and synergistically increases intestinal inflammation when absent. Increased numbers of TdT-mediated dUTP nick end labelling (TUNEL)<sup>+</sup> cells in *Atg16l1/Xbp1<sup>AIEC</sup>* and *Atg7/Xbp1<sup>AIEC</sup>* mice (Extended Data Fig. 6a, b) correlated with enteritis severity in double-mutant, in contrast to *Xbp1<sup>AIEC</sup>* mice (Extended Data Fig. 6b), and, as demonstrated for *Atg16l1/Xbp1<sup>AIEC</sup>* mice, concomitantly with increasing age (Extended Data Fig. 6c and Supplementary Video 1). Silencing of *Atg16l1* in *shXbp1* MODE-K cells significantly increased the proportion of apoptotic cells *in vitro* (Fig. 3a and Extended Data Fig. 6d), indicating that increased apoptosis could function as an initial event in intestinal inflammation. Furthermore, intestinal-epithelial-cell-associated NF- $\kappa$ B activation, a critical player in intestinal inflammation<sup>26</sup>, was absent in primary intestinal epithelial cells from wild-type mice and gradually increased in *Atg16l1<sup>AIEC</sup>*, *Xbp1<sup>AIEC</sup>* and *Atg16l1/Xbp1<sup>AIEC</sup>* mice (Fig. 3b and Extended Data Fig. 6e, f). *shXbp1* MODE-K cells exhibited increased NF- $\kappa$ B activation when stimulated with tumour-necrosis factor (TNF) (Fig. 3c and Extended Data Fig. 6g, h) or Toll-like receptor ligands (data not shown) relative to shCtrl MODE-K cells, demonstrating increased sensitivity of XBP1-deficient intestinal epithelial cells to inflammatory and environmental stimuli. Inhibition of NF- $\kappa$ B by treatment with BAY11-7082 decreased intestinal epithelial cell death in *Xbp1<sup>T-AIEC</sup>* mice (Fig. 3d and Extended Data Fig. 6i) and protected from enteritis in both *Xbp1<sup>T-AIEC</sup>* and *Xbp1<sup>AIEC</sup>* mice (Fig. 3e and Extended Data Fig. 6j) relative to vehicle-treated mice. A progressive increase in intestinal epithelial cells of total and phosphorylated IRE1 $\alpha$  (encoded by *Ern1*), the sensor of ER stress upstream of XBP1 and known to control NF- $\kappa$ B<sup>27</sup>, was observed in *Atg16l1<sup>AIEC</sup>*, *Xbp1<sup>AIEC</sup>* and *Atg16l1/Xbp1<sup>AIEC</sup>* mice in comparison to wild-type mice (Fig. 3f) and mirrored the escalating elevations in NF- $\kappa$ B (Fig. 3b and Extended Data Fig. 6e, f). Indeed, increased NF- $\kappa$ B activity and ileitis were governed by IRE1 $\alpha$ , as the increased epithelial NF- $\kappa$ B phosphorylation observed in *Xbp1<sup>AIEC</sup>* mice was abrogated in *Ern1/Xbp1<sup>AIEC</sup>* mice (Fig. 3g), *Ern1* co-silencing of *shXbp1* MODE-K cells abolished the increased expression of the NF- $\kappa$ B target gene *Nfkb1a* in response to TNF stimulation relative to shCtrl MODE-K cells (Extended Data Fig. 6k), and enteritis was diminished in *Ern1/Xbp1<sup>AIEC</sup>* compared to *Xbp1<sup>AIEC</sup>* mice (Fig. 3h). Enteritis was also reversed by germline deletion of *Tnfrsf1a* in *Xbp1<sup>AIEC</sup>* mice (Fig. 3i), and by re-derivation of *Xbp1<sup>AIEC</sup>* mice into a germ-free environment (Fig. 3j), which was associated with reduced p-IkB $\alpha$  immunoreactivity in *Xbp1*-deficient intestinal epithelial cells compared to mice housed under specific pathogen-free (SPF) conditions (Fig. 3k). These studies together demonstrate that enteritis in this model is driven by TNF—the cytokine targeted by the most potent therapeutics of human Crohn's disease<sup>25</sup>—and microbes in a pathway that derives from intestinal epithelial cell death and IRE1 $\alpha$ -dependent activation of NF- $\kappa$ B, with ATG16L1-dependent autophagy serving to restrain the inflammatory nature of the latter, probably through the removal of IRE1 $\alpha$ .

Accordingly, treatment of *Xbp1<sup>T-AIEC</sup>* mice with the mTOR inhibitor rapamycin<sup>10</sup> induced autophagy primarily in crypts (Extended Data Fig. 7a, b), diminished intestinal-epithelial-cell-associated NF- $\kappa$ B activation and number of TUNEL<sup>+</sup> intestinal epithelial cells in *Xbp1<sup>AIEC</sup>* mice (Fig. 4a and Extended Data Fig. 7c–e), and markedly reduced the severity of enteritis (Fig. 4b). These beneficial consequences of rapamycin in the setting of unabated ER stress in intestinal epithelial cells were not



**Figure 4 | ER stress-induced enteritis originates from Paneth cells and is alleviated through autophagy induction.** **a**, Immunoblot of primary intestinal epithelial cell scrapings from mice treated with or without rapamycin for 14 consecutive days ( $n = 3$ ). **b**, Enteritis histology score for experiment as in **a** ( $n = 4$ ; median shown; Mann–Whitney *U*-test). **c**, Representative images of EYFP-*Rosa26*/D6-*cre*<sup>+/−</sup> reporter mice and EYFP-*Rosa26* (controls). Co-localization of *Defa6* Cre-driven EYFP expression (yellow) with lysozyme-expressing Paneth cells (red;  $n = 3$ ). DAPI, blue; scale bars, 50  $\mu$ m. **d**, Immunoblots of crypt intestinal epithelial cells from *Xbp1<sup>APC</sup>* and wild-type controls ( $n = 2$ ). **e**, **f**, Representative confocal images of lysozyme (green) expressing Paneth cells (**e**) with quantification of crypts with indicated number of lysozyme<sup>+</sup> granulated dots in **f** ( $n = 5$ ; unpaired Student's *t*-test; mean  $\pm$  s.e.m.). DAPI, blue; scale bars, 10  $\mu$ m. **g**, Immunohistochemistry for p-eIF2 $\alpha$  ( $n = 3$ ). Scale bars, 20  $\mu$ m. **h**, Representative haematoxylin and eosin images of *Xbp1<sup>APC</sup>* and wild-type mice scored in **i**. Scale bars, 50  $\mu$ m. **i**, Enteritis scoring in *Xbp1<sup>fl/fl</sup>* (WT<sup>fl</sup>), *Defa6-cre*<sup>+</sup> (WT<sup>cre</sup>) and *Defa6-cre*<sup>+</sup>; *Xbp1<sup>fl/fl</sup>* (*Xbp1<sup>APC</sup>*) mice ( $n = 21/26/29$ ; median shown; Kruskal–Wallis with post-hoc Holm's-corrected Mann–Whitney *U*-test). Results represent two (**b**, **d**) independent experiments. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

observed in *Atg7/Xbp1<sup>AIEC</sup>* and *Atg16l1/Xbp1<sup>AIEC</sup>* mice (Extended Data Fig. 7e–i), demonstrating that these effects required intact autophagy within intestinal epithelial cells.

The spatial convergence of the consequences of hypomorphic UPR and autophagy function in Paneth cells prompted us to test the hypothesis that these pathways were interdependent in this cell type. We developed a defensin 6 alpha promoter<sup>28</sup>-driven *cre* transgenic line (D6-*cre*), and confirmed the exclusive activity of Cre recombinase in lysozyme<sup>+</sup> Paneth cells (Fig. 4c). Paneth-cell-specific deletion of *Xbp1* in D6-*cre*; *Xbp1<sup>fl/fl</sup>* (*Xbp1<sup>APC</sup>*) mice resulted in autophagy activation (Fig. 4d and

Extended Data Fig. 8a, b) and structural defects in granule morphology in Paneth cells (Fig. 4e, f). Crypts of *Xbp1*<sup>ΔPC</sup> compared to wild-type mice exhibited increased GRP78 expression (Fig. 4d and Extended Data Fig. 8b) and p-eIF2α immunoreactivity (Fig. 4g), accompanied by increased conversion of LC3-I/II and reduced p62 levels (Fig. 4d and Extended Data Fig. 8b), demonstrating ER stress and autophagy induction. Notably, 75% of *Xbp1*<sup>ΔPC</sup> mice developed spontaneous enteritis (Fig. 4h, i) that shared the histological inflammatory characteristics observed in *Xbp1*<sup>ΔIEC</sup> mice (Fig. 2b). *Xbp1*<sup>ΔPC</sup> mice exhibited increased cell death in crypts (Extended Data Fig. 9a, b) and intestinal epithelial cell turnover (Extended Data Fig. 9c–f), whereas goblet cells remained unaffected (Extended Data Fig. 9g, h). Notably, *Atg7/Xbp1*<sup>ΔPC</sup> mice exhibited transmural disease as early as 8 weeks of age (Extended Data Fig. 9i). We conclude that deletion of *Xbp1* specifically in Paneth cells results in unresolved ER stress, UPR activation and induction of autophagy, which can serve as a nidus for the emergence of spontaneous intestinal inflammation that evolves into transmural disease in the absence of the compensation provided by autophagy.

Our studies thus support a mechanistic model (Extended Data Fig. 10a–c) for how *ATG16L1*-associated genetic risk may convert into a disease phenotype. The competence of the UPR probably sets the threshold for susceptibility of the host with hypofunctional autophagy to interacting genetic and environmental factors capable of inducing inflammation. Consistent with our model, patients with Crohn's disease carrying the *ATG16L1*<sup>T300A</sup> risk variant, which impairs autophagosome formation<sup>29</sup>, frequently exhibit ER stress in their Paneth cells, in contrast to those harbouring the normal variant<sup>30</sup>. Finally, our studies also unequivocally establish that these inflammatory susceptibilities can emerge directly from highly secretory Paneth cells and suggest that small intestinal Crohn's disease may be a specific disorder of this cell type.

## METHODS SUMMARY

Villin (*V*)-*cre*<sup>+</sup>; *Xbp1*<sup>fl/fl</sup> (*Xbp1*<sup>ΔIEC</sup>)<sup>4</sup>, *V*-*cre*ER<sup>T2+</sup>; *Xbp1*<sup>fl/fl</sup> (*Xbp1*<sup>T-ΔIEC</sup>)<sup>4</sup>, *Atg7*<sup>fl/fl</sup> (ref. 14), *Atg16l1*<sup>fl/fl</sup> (generated as detailed in Methods) and *Gadd34*<sup>−/−</sup> mice<sup>16</sup>, all backcrossed >8 generations onto B6, and *GFP-LC3* (ref. 12), *Ern1*<sup>fl/fl</sup> and *Tnfrsf1a*<sup>−/−</sup> mice were used to generate experimental mouse strains. Experiments were performed with sex- and age-matched animals using littermates where technically possible. Generation of Paneth-cell-specific *Defa6-cre*; *Xbp1*<sup>fl/fl</sup> (*Xbp1*<sup>ΔPC</sup>) mice is described in Methods. Murine norovirus (MNV) status is reported in Extended Data Fig. 9j. Stable clones of the *Xbp1*-silenced (*shXbp1*) small intestinal epithelial cell line MODE-K were generated via a lentiviral shRNA vector and selected by hygromycin<sup>4</sup>. The sources of antibodies and reagents are listed in the Methods. Immunoblot, chromatin immunoprecipitation, qRT-PCR, histological analysis, transmission electron microscopy and confocal microscopy techniques are described in Methods. Statistical significance was calculated using an unpaired two-tailed Student's *t*-test or a Mann-Whitney *U*-test and considered significant at *P* < 0.05. In experiments where more than two groups were compared, Kruskal–Wallis test followed by Mann–Whitney *U* and post-hoc Bonferroni Holm's correction or one-way ANOVA/Bonferroni were performed as appropriate. Data were analysed using GraphPad Prism software.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 June; accepted 22 August 2013.

Published online 2 October 2013.

- Jostins, L. *et al.* Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012).
- Cadwell, K. *et al.* A key role for autophagy and the autophagy gene *Atg16l1* in mouse and human intestinal Paneth cells. *Nature* **456**, 259–263 (2008).
- Cadwell, K. *et al.* Virus-plus-susceptibility gene interaction determines Crohn's disease gene *Atg16l1* phenotypes in intestine. *Cell* **141**, 1135–1145 (2010).
- Kaser, A. *et al.* XBP1 links ER stress to intestinal inflammation and confers genetic risk for human inflammatory bowel disease. *Cell* **134**, 743–756 (2008).
- Tréton, X. *et al.* Altered endoplasmic reticulum stress affects translation in inactive colon tissue from patients with ulcerative colitis. *Gastroenterology* **141**, 1024–1035 (2011).
- Kobayashi, K. S. *et al.* Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. *Science* **307**, 731–734 (2005).
- Zhao, F. *et al.* Disruption of Paneth and goblet cell homeostasis and increased endoplasmic reticulum stress in *Ag2*<sup>−/−</sup> mice. *Dev. Biol.* **338**, 270–279 (2010).
- Wehkamp, J. *et al.* Reduced Paneth cell α-defensins in ileal Crohn's disease. *Proc. Natl Acad. Sci. USA* **102**, 18129–18134 (2005).

- Maloy, K. J. & Powrie, F. Intestinal homeostasis and its breakdown in inflammatory bowel disease. *Nature* **474**, 298–306 (2011).
- Levine, B., Mizushima, N. & Virgin, H. W. Autophagy in immunity and inflammation. *Nature* **469**, 323–335 (2011).
- Sarkar, S., Ravikumar, B. & Rubinsztein, D. C. Autophagic clearance of aggregate-prone proteins associated with neurodegeneration. *Methods Enzymol.* **453**, 83–110 (2009).
- Mizushima, N., Yamamoto, A., Matsui, M., Yoshimori, T. & Ohsumi, Y. *In vivo* analysis of autophagy in response to nutrient starvation using transgenic mice expressing a fluorescent autophagosome marker. *Mol. Biol. Cell* **15**, 1101–1111 (2004).
- Rouschop, K. M. *et al.* The unfolded protein response protects human tumor cells during hypoxia through regulation of the autophagy genes MAP1LC3B and ATG5. *J. Clin. Invest.* **120**, 127–141 (2010).
- Komatsu, M. *et al.* Impairment of starvation-induced and constitutive autophagy in *Atg7*-deficient mice. *J. Cell Biol.* **169**, 425–434 (2005).
- Boyce, M. *et al.* A selective inhibitor of eIF2α dephosphorylation protects cells from ER stress. *Science* **307**, 935–939 (2005).
- Novoa, I. *et al.* Stress-induced gene expression requires programmed recovery from translational repression. *EMBO J.* **22**, 1180–1187 (2003).
- Hampe, J. *et al.* A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in *ATG16L1*. *Nature Genet.* **39**, 207–211 (2007).
- Cleynen, I. *et al.* Genetic factors conferring an increased susceptibility to develop Crohn's disease also influence disease phenotype: results from the IBDchip European Project. *Gut* <http://dx.doi.org/10.1136/gutjnl-2011-300777> (2012).
- Fujita, N. *et al.* The Atg16L complex specifies the site of LC3 lipidation for membrane biogenesis in autophagy. *Mol. Biol. Cell* **19**, 2092–2100 (2008).
- Mizushima, N. *et al.* Mouse Apg16L, a novel WD-repeat protein, targets to the autophagic isolation membrane with the Apg12–Apg5 conjugate. *J. Cell Sci.* **116**, 1679–1688 (2003).
- Fujita, N. *et al.* Differential involvement of Atg16L1 in Crohn disease and canonical autophagy: analysis of the organization of the Atg16L1 complex in fibroblasts. *J. Biol. Chem.* **284**, 32602–32609 (2009).
- Yang, L., Li, P., Fu, S., Calay, E. S. & Hotamisligil, G. S. Defective hepatic autophagy in obesity promotes ER stress and causes insulin resistance. *Cell Metab.* **11**, 467–478 (2010).
- Bertolotti, A. *et al.* Increased sensitivity to dextran sodium sulfate colitis in IRE1 β-deficient mice. *J. Clin. Invest.* **107**, 585–593 (2001).
- Cao, S. S. *et al.* The unfolded protein response and chemical chaperones reduce protein misfolding and colitis in mice. *Gastroenterology* **144**, 989–1000 (2013).
- Kaser, A., Zeissig, S. & Blumberg, R. S. Inflammatory bowel disease. *Annu. Rev. Immunol.* **28**, 573–621 (2010).
- Rogler, G. *et al.* Nuclear factor κB is activated in macrophages and epithelial cells of inflamed intestinal mucosa. *Gastroenterology* **115**, 357–369 (1998).
- Kaneko, M., Niinuma, Y. & Nomura, Y. Activation signal of nuclear factor-κB in response to endoplasmic reticulum stress is transduced via IRE1 and tumor necrosis factor receptor-associated factor 2. *Biol. Pharm. Bull.* **26**, 931–935 (2003).
- Vaishnava, S., Behrendt, C. L., Ismail, A. S., Eckmann, L. & Hooper, L. V. Paneth cells directly sense gut commensals and maintain homeostasis at the intestinal host-microbial interface. *Proc. Natl Acad. Sci. USA* **105**, 20858–20863 (2008).
- Kuballa, P., Huett, A., Rioux, J. D., Daly, M. J. & Xavier, R. J. Impaired autophagy of an intracellular pathogen induced by a Crohn's disease associated *ATG16L1* variant. *PLoS ONE* **3**, e3391 (2008).
- Deuring, J. J. *et al.* Genomic *ATG16L1* risk allele-restricted Paneth cell ER stress in quiescent Crohn's disease. *Gut* <http://dx.doi.org/10.1136/gutjnl-2012-303527> (2013).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank L. Glimcher, A. Goldberg, J. Yuan, M. Parkes, A. Franke, H. Tilg, M. Pasparakis, K. Vantis, A.-H. Lee and C. L. Bevins for discussion of the project, are grateful to J. Gordon, L. Hooper and K. Rajewsky for providing critical reagents, and thank O. Will for initial handling of the *Atg16l1* colony and help with DSS colitis. A. Kaser began work for this study at the Department of Internal Medicine II, Innsbruck Medical University, A-6020 Innsbruck, Austria. This work was supported by NIH grants DK044319, DK051362, DK053056, DK088199, the Harvard Digestive Diseases Center (HDDC) (DK0034854) (R.S.B.); the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC Grant agreement no. 260961 (A.K.); the National Institute for Health Research Cambridge Biomedical Research Centre (A.K.); the Austrian Science Fund and Ministry of Science P21530-B18 and START Y446-B18 (A.K.); the Addenbrooke's Charitable Trust (A.K. and L.N.); BMBF NGFN Animal Model grant (P.R.), the DFG Cluster of Excellence Inflammation at Interfaces (S.S. and P.R.); EU SysmedIBD grant (P.R.), the Hans-Dietrich Bruhn Memorial Foundation (R.B.); DFG grants RO2994/5-1 (P.R.) and SFB 877 project B9 (P.R. and S.S.); fellowships from Inflammatory Bowel Disease Working Group (M.F.T.), Crohn's and Colitis Foundation of America (M.B.F.), European Crohn's and Colitis Organization (T.E.A.), Crohn's in Childhood Research Association (A.K. and L.N.), National Research Foundation of Korea funded by the Korean government KRF-2008-357-E00022 no. 2011-0009018 (H.-J.K.).

**Author Contributions** T.E.A., M.F.T., L.N. and H.-J.K. performed most experiments, together with J.B., E.M.-N., M.T., S.H., M.B.F., S.B.-B., T.R., R.B. and M.-N.K. J.L.C. helped prepare the manuscript. S.J.H. and J.H. contributed electron microscopic analysis. A.C. provided expertise in autophagy assessment, and R.L. in histology of Paneth cells. S.S. and P.R. designed, generated and analysed an essential mouse strain. K.K., T.I. and S.J.M. provided an essential mouse strain. J.N.G. assessed intestinal inflammation. A.K. and R.S.B. devised and coordinated the project, and together with T.E.A. and M.F.T. wrote the manuscript and designed the experiments.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.S.B. ([rblumberg@partners.org](mailto:rblumberg@partners.org)) or A.K. ([ak729@cam.ac.uk](mailto:ak729@cam.ac.uk)).



## METHODS

**Mice.** *Xbp1<sup>fl/fl</sup>* mice were backcrossed >8 generations onto B6 and crossed with *V-crc-ER<sup>T2+</sup>* (B6) mice provided by N. Davidson and S. Robine to generate mice with tamoxifen-inducible *Xbp1* deletion in the intestinal epithelium (*Xbp1<sup>T-AIEC</sup>*), and with *V-crc* (B6) mice to generate mice with constitutively active *Xbp1* deletion in the intestinal epithelium (*Xbp1<sup>AIEC</sup>*). *V-crc-ER<sup>T2+</sup>* recombinase was activated by 3 or 5 daily intraperitoneal administrations of 1 mg tamoxifen (MP Biomedicals) as indicated in figure legends. *V-crc<sup>+</sup>;Xbp1<sup>fl/fl</sup>* (*Xbp1<sup>AIEC</sup>*) mice were crossed with *Atg7<sup>fl/fl</sup>* mice<sup>14</sup> provided by M. Komatsu to obtain *V-crc<sup>+</sup>;Atg7<sup>fl/fl</sup>;Xbp1<sup>fl/fl</sup>* (*Atg7/Xbp1<sup>AIEC</sup>*) mice. Mice with a floxed *Atg16l1* allele were generated in collaboration with GenOway. Briefly, a proximal *loxP* site was introduced within the promoter region of the *Atg16l1* gene upstream of exon 1, a distal *loxP* site was introduced with an FRT flanked neomycin selection cassette within intron 1. The resultant mouse line was bred with deleter-mice constitutively expressing Flp recombinase to remove the neomycin selection cassette, creating an *Atg16l1<sup>fl/+</sup>* mouse in which *Atg16l1* exon 1 was flanked by two *loxP* sites (Extended Data Fig. 4g). After backcrossing onto B6, these mice were crossed with *V-crc<sup>+</sup>* mice<sup>31</sup> resulting in *V-crc;Atg16l1<sup>fl/fl</sup>* mice with intestinal-epithelial-cell-specific *Atg16l1* deletion (*Atg16l1<sup>AIEC</sup>*). *Atg16l1<sup>fl/fl</sup>* mice were crossed with *Xbp1<sup>AIEC</sup>* mice to develop *V-crc<sup>+</sup>;Atg16l1<sup>fl/fl</sup>;Xbp1<sup>fl/fl</sup>* (*Atg16l1/Xbp1<sup>AIEC</sup>*) mice. GFP-LC3 transgenic mice<sup>12</sup> (gift of N. Mizushima) were crossed with *V-crc-ER<sup>T2+</sup>;Xbp1<sup>fl/fl</sup>* to generate GFP-LC3;*V-crc-ER<sup>T2+</sup>;Xbp1<sup>fl/fl</sup>* mice (GFP-LC3;*Xbp1<sup>T-AIEC</sup>*). For the generation of Paneth-cell-specific *Defa6-crc;Xbp1<sup>fl/fl</sup>* mice, a 1.1-kb cDNA fragment encoding improved Cre (iCre)<sup>32</sup> recombinase was subcloned downstream of nucleotides -6,500 to +34 of mouse cryptdin-2 gene (*Defa6*) in the BamHI site of the pCR2-TAG-hGH plasmid<sup>33,34</sup> to replace the DNA fragment containing the simian virus 40 large antigen (SV40). A linearized 10.2-kb fragment containing the *Defa6* promoter, iCre and hGH (*Defa6-iCre-hGH*) was removed by EcoRI digestion, agarose gel-electrophoresed and purified with the Qiaex Gel Extraction kit (Qiagen), and used for pronuclear injection of B6 mice. Six founders from 22 live-born mice were identified by screening tail DNA using iCre-specific primers, and two lines were further characterized and crossed to *Xbp1<sup>fl/fl</sup>*, *Atg7<sup>fl/fl</sup>* and *EYFP-Rosa26* reporter mice<sup>35</sup>, provided by K. Rajewsky. *Gadd34<sup>-/-</sup>* mice<sup>16</sup> were crossed with *V-crc;Xbp1<sup>fl/fl</sup>* mice to generate *V-crc;Xbp1<sup>fl/fl</sup>;Gadd34<sup>-/-</sup>* (*Xbp1<sup>AIEC</sup>;Gadd34<sup>-/-</sup>*) mice. *Ern1<sup>fl/fl</sup>* (ref. 36) and *Tnfrsf1a<sup>-/-</sup>* (Jackson) mice were crossed with *V-crc;Xbp1<sup>fl/fl</sup>* mice to generate *V-crc;Ern1<sup>fl/fl</sup>;Xbp1<sup>fl/fl</sup>* (*Ern1/Xbp1<sup>AIEC</sup>*) and *V-crc;Xbp1<sup>fl/fl</sup>;Tnfrsf1a<sup>-/-</sup>* (*Xbp1<sup>AIEC</sup>;Tnfrsf1a<sup>-/-</sup>*) mice, respectively. Cre transgenes were maintained in the hemizygous state in all experimental strains with a floxed allele to generate littermate controls. *Xbp1<sup>AIEC</sup>* mice were re-derived in a germ-free environment and housed in sterile isolators at the Taconic Farms breeding facility. Tail or ear biopsy genomic DNA was used for genotyping of respective mouse strains as described previously<sup>4</sup>. Primer sequences are available on request. Mice were housed in specific pathogen free (SPF) barrier facilities at Harvard Medical School (*Xbp1<sup>AIEC</sup>*, *Xbp1<sup>T-AIEC</sup>*, *Atg7<sup>AIEC</sup>*, *Atg7/Xbp1<sup>AIEC</sup>*, *Atg7/Xbp1<sup>APC</sup>*, *Xbp1<sup>APC</sup>*, GFP-LC3;*Xbp1<sup>AIEC</sup>*, *Ern1/Xbp1<sup>AIEC</sup>* mice and their respective controls), University of Cambridge (*Xbp1<sup>AIEC</sup>*, *Xbp1<sup>T-AIEC</sup>*, *Atg16l1<sup>AIEC</sup>*, *Atg16l1/Xbp1<sup>AIEC</sup>*, *Xbp1<sup>AIEC</sup>/Gadd34<sup>-/-</sup>*, *Ern1/Xbp1<sup>AIEC</sup>* mice and their respective controls), Innsbruck Medical University (*Xbp1<sup>AIEC</sup>*, *Xbp1<sup>T-AIEC</sup>*, *Atg16l1<sup>AIEC</sup>*, *Atg16l1/Xbp1<sup>AIEC</sup>*, *Xbp1<sup>AIEC</sup>/Tnfrsf1a<sup>-/-</sup>* mice and their respective controls), and Christian-Albrechts-Universität zu Kiel (*Atg16l1<sup>AIEC</sup>* mice and their respective controls). Colonies maintained at Boston and Innsbruck were murine norovirus (MNV) positive by Taqman qRT-PCR (Extended Data Fig. 9j). *Xbp1<sup>AIEC</sup>*, *Atg16l1<sup>AIEC</sup>*, *Ern1<sup>AIEC</sup>* and their associated double-mutant strains were re-derived from the Innsbruck colony into the MNV-free enhanced barrier Cambridge facility, and colonies confirmed MNV-negative by PCR (Extended Data Fig. 9j) and serology (data not shown), as were *Atg16l1<sup>AIEC</sup>* mice held at the Kiel facility. MNV Taqman qRT-PCR was performed as described<sup>37</sup>. The phenotype of single- and double-mutant colonies that had been re-derived from the MNV<sup>+</sup> Innsbruck facility into the MNV<sup>-</sup> Cambridge facility were indistinguishable, in particular relating to qualitative and quantitative measures of enteritis and the reciprocal induction of autophagy and ER stress. Mice were handled and all experiments performed in accordance with institutional guidelines and with the approval of the relevant authorities. 8–10-week-old mice were used for all experiments unless stated otherwise in the figure legend, and were randomly allocated into treatment groups.

**Antibodies and reagents.** The following antibodies and reagents were used for immunoblotting. Sigma Aldrich: anti-LC3B (L7543); Cell Signaling Technology: anti-β-actin (4970; 13E5) anti-GAPDH (2118; 14C10), anti-eIF2α (9722), anti-phospho-eIF2α (3597; 119A11), anti-PERK (3192; C33E10), anti-phospho-PERK (3179; 16F8), anti-JNK (9252), anti-phospho-JNK (4668; 81E11), anti-ATG5 (8540; D1G9), anti-beclin 1 (3495; D4OC5), anti-ATG7 (8558; D12B11), anti-CHOP (5554; D46F1), anti-ATG12 (4180; D88H11), anti-p62 (5114), anti-IKK1 (2682), anti-IKK2 (2370; 2C8), anti-phospho-IKK1/2 (2697; 16A6), anti-IRE1α (3294; 14C10), anti-phospho-NF-κB p65 (3033; 93H1), anti-NF-κB p65 (4764; C22B4) and anti-rabbit/mouse HRP antibodies (7074, 7076); Abcam: anti-phospho-IRE1α (48187);

MBL: anti-ATG16L1 (M150-3; 1F12); Stressgen: anti-haem-oxygenase-1 (ADI-SPA-895); Novus Biologicals: anti-GRP78 (NBP1-06274); Santa Cruz Biotechnology: anti-ATF4 (sc-200; C20). Immunoprecipitation antibody: anti-IRE1α (Santa Cruz Biotechnology, 20790; H190). Immunohistochemistry antibodies: Santa Cruz Biotechnology: anti-lysozyme (27958; C19); MBL: anti-ATG16L1 (M150-3; 1F12); Cell Signaling Technology: anti-ATG16L1 (8089; D6D5), anti-phospho-IκBα (2859; 14D4), anti-phospho-eIF2α (3597; 119A11); Abcam: anti-Ki67 (15580), anti-GRP78 (21685); Progen: anti-p62 (GP62-C); BD Bioscience: anti-BrdU (551321).

The following reagents were used: TNF (Peprotech, 315-01A), bafilomycin A1 (Sigma Aldrich, B1793), rapamycin (LC Laboratories, R-5000), JNK inhibitor (Sigma Aldrich, SP600125), NF-κB inhibitor (Calbiochem, BAY11-7082) and salubrin (Alexis Biochemicals, ALX-270-428) were dissolved in DMSO as recommended. *N*-acetyl cysteine (Sigma Aldrich, A9165) and glutathione (Calbiochem, NOV3541) were used at final concentration of 1 mM. Ambion siRNA for *Atg16l1* (94892, sense 5'-GAACUGUUAGGGAAGAUCAATT-3', antisense 5'-UGAUCU UCCCUAACAGUCCCA-3'), *Perk* (65405, sense 5'-CCCAGUAUCUACAGAUUUTT-3', antisense 5'-AAAUCUGUUAGAUUACGGGAT-3'; 65406, sense 5'-CGAAGAAUACAGUAAUGGUTT-3', antisense 5'-ACCAUACUGUAU CUUCGTG-3'), *Gadd34* (70230, sense 5'-CCAUAGCUCGGGAUACAATT-3', antisense 5'-UUGUAUCCCGGAGCUAUGGAA-3'; 70231, sense 5'-AGACAAC AGCAUUCGGAUUTT-3', antisense 5'-AUCCGAAUCGUGUUGUCUTC-3'), *Ern1* (95857, sense 5'-GUUUGACCCUGGACUACAAATT-3', antisense 5'-U UUGAGUCCAGGGUCAAACCTT-3'; 95858, sense 5'-GGAUGUAAGUGACCG AAUATT-3', antisense 5'-UAUUCGGGUCACUUAUACUCCCTG-3'; 95859, sense 5'-GCUCGUGAAUUGAUAGAGATT-3', antisense 5'-UCUCUAUCAAUUAU CAGCAA-3') and scrambled control were used at a final concentration of 10 μM.

**Chromatin immunoprecipitation.** ChIP with anti-ATF4 (Santa Cruz Biotechnology) and control IgG rabbit antibody was performed in *Xbp1* and control silenced MODE-K cells according to ChIP protocol by Agilent. To determine the presence of ATF4 binding sites in the *Atg7* promoter, a 4-kb region proximal to the transcription start site identified with the Eukaryotic Promoter Database Primers was analysed using MatInspector (Genomatix). Immunoprecipitated DNA was subjected to quantitative PCR (qPCR) to determine enrichment of ATF4 binding to respective promoters and results were normalized to input chromatin DNA. Primers used for qPCR were as follows for *Map1LC3b* (ref. 13), for *Atg7* forward 5'-GCGCTTCGCGTTTGTGTGG-3' and reverse 5'-CTGCTCCGAACAC GGCTT-3'.

**Salubrin, rapamycin and BAY11-7082 treatment in vivo.** Salubrin (1 mg kg<sup>-1</sup> d<sup>-1</sup>), rapamycin (1.5 mg kg<sup>-1</sup> d<sup>-1</sup>) or vehicle (DMSO) was administered intraperitoneally (i.p.) 24 h before the first tamoxifen administration to GFP-LC3;*V-crc-ER<sup>T2</sup>;Xbp1<sup>fl/fl</sup>* (GFP-LC3;wild type) and GFP-LC3;*Xbp1<sup>T-AIEC</sup>* mice. 3-day treatment was used for evaluation of accumulation of GFP-LC3 punctae in the intestinal epithelium, whereas a 5-day combined tamoxifen and salubrin treatment followed by two daily salubrin injections was used in experiments with enteritis assessment as an end point (Extended Data Figs 2a and 3g). To assess the effects of rapamycin on ER stress-induced intestinal inflammation in *XBP1* deficiency, *V-crc;Xbp1<sup>fl/fl</sup>* (*Xbp1<sup>AIEC</sup>*), *V-crc;Atg16l1<sup>fl/fl</sup>;Xbp1<sup>fl/fl</sup>* (*Atg16l1/Xbp1<sup>AIEC</sup>*) or *V-crc;Atg7<sup>fl/fl</sup>;Xbp1<sup>fl/fl</sup>* (*Atg7/Xbp1<sup>AIEC</sup>*) and the respective control mice were treated with rapamycin or vehicle for 14 consecutive days i.p. and inflammation was evaluated. BAY11-7082<sup>38</sup> or vehicle (DMSO) was administered i.p. every other day at 5 mg kg<sup>-1</sup> for 14 consecutive days in *Xbp1<sup>AIEC</sup>* mice, or for 5 consecutive days at 20 mg kg<sup>-1</sup> in *Xbp1<sup>T-AIEC</sup>* mice concomitant with i.p. tamoxifen.

**Transmission electron microscopy.** Small intestinal tissue from mice was handled by standard methods to be fixed with 1.25% glutaraldehyde, 4% formaldehyde in 0.1 M cacodylate buffer at pH 7.4 at room temperature for electron microscopy. The detailed procedures for electron microscopy was previously described<sup>39</sup> and the tissue was observed with a JEOL 1400 transmission electron microscope at 120 kV operating voltage. For quantification of autophagy, number of autophagic vacuoles was manually counted by a TEM expert (J.H.) blinded to sample identity in 10 consecutive Paneth cells per sample. ImageJ software was used to measure average size of autophagic vacuoles.

**Histology.** Formalin-fixed and paraffin-embedded intestinal tissue was sectioned and stained with haematoxylin and eosin as previously described<sup>4</sup>. A semi-quantitative composite scoring system was used for the assessment of spontaneous intestinal inflammation, computed as a sum of five histological subscores, multiplied by a factor based on the extent of the inflammation. Histological subscores (for each parameter: 0, absent; 1, mild; 2, moderate; 3, severe): mononuclear cell infiltrate (0–3), crypt hyperplasia (0–3), epithelial injury/erosion (0–3), polymorphonuclear cell infiltrates (0–3) and transmural inflammation (0, absent; 1, submucosal; 2, one focus extending into muscularis and serosa; 3 up to five foci extending into muscularis and serosa; 4, diffuse). Extent factor was derived according to the fraction of bowel length involved by inflammation: 1, <10%; 2, 10–25%; 3, 25–50%; and



4, >50%. Ileal inflammation was assessed by an expert gastrointestinal pathologist (J.N.G.) who was blinded to the genotype and experimental conditions of the samples. No spontaneous colonic inflammation was detected in any of the reported genotypes.

**Reactive oxygen species, cell death detection by flow cytometry and NF- $\kappa$ B activity assays.** Stable clones of the *Xbp1*-silenced (*shXbp1*) small intestinal epithelial cell line MODE-K were generated via a lentiviral shRNA vector and selected by hygromycin, as previously described<sup>4</sup>. To evaluate oxidative stress, *Xbp1*- and control-silenced MODE-K cells<sup>40</sup> were incubated with 5  $\mu$ M 5-(and-6)-chloromethyl-2',7'-dichlorodihydrofluorescein diacetate, acetyl ester (CM-H2DCFDA) (Molecular Probes) for 30 min<sup>41</sup>. After washing with PBS, cells were further incubated with complete medium for 2 h. Reactive oxygen species generation was determined using flow cytometry. To evaluate cell death, *shXbp1* and *shCtrl* cells were co-silenced for *Atg16l1* using siRNA or scrambled control (Ambion). After 4 days, cells were collected and stained for annexin V (Biolegend) in staining buffer (Biolegend) and mode of cell death was determined by flow cytometry after addition of propidium iodide (PI). To assess NF- $\kappa$ B signalling pathway activation, *Xbp1*- or control-silenced MODE-K cells were stimulated with 50 ng ml<sup>-1</sup> TNF for indicated periods of time, followed by immunoblotting (using NE-PER (Thermo Scientific) isolated cytoplasmic extracts), qRT-PCR, and chemiluminescent detection of NF- $\kappa$ B consensus sequence binding activity (in nuclear extracts isolated with NE-PER) with the NF- $\kappa$ B p65 transcription factor assay kit (Thermo Scientific). *Ern1* or scrambled siRNA (Ambion) was used for co-silencing as indicated.

**Immunohistochemistry, BrdU and TUNEL labelling.** Formalin-fixed paraffin-embedded sections were stained according to standard immunohistochemistry protocols and manufacturer's recommendations as described previously<sup>4</sup>. Cell death was assessed by TdT-mediated dUTP nick end labelling (TUNEL) of formalin-fixed paraffin-embedded slides of the respective genotypes using the TUNEL cell death detection kit (Roche). Entire slides were analysed for TUNEL<sup>+</sup> cells and numbers normalized to intestinal length on the slide. Proliferation of the intestinal epithelium was assessed after a 24-h pulse with 5-bromodeoxyuridine (BrdU; BD Pharmingen) and incorporated BrdU was detected by the BrdU *in situ* detection kit (BD Pharmingen). BrdU<sup>+</sup> nuclei per total intestinal epithelial cells along the crypt villus axis are shown. Toluidine blue staining and Periodic acid-Schiff (PAS) reaction was performed according to standard protocols.

**Confocal microscopy for detection of GFP-LC3 and EYFP.** For detection of GFP-LC3 or EYFP, mice were euthanized, followed by transcardiac perfusion with PBS (2–3 min) and 3.7% formaldehyde (3–4 min). Small intestine was dissected and promptly washed with PBS. The tissue was fixed in formalin for an additional 12–18 h. Fixed tissue was embedded in OCT and sectioned on a cryotome into 5  $\mu$ m sections. Slides were washed with PBS and mounted with Prolong Gold Antifade reagent with DAPI (Invitrogen). Images of the sections were collected using Olympus semi-confocal system. MetaMorph software was used for image analysis. For the detection of autophagosome formation *in vitro*, *Xbp1* and control silenced MODE-K cells were transfected with a GFP-LC3 plasmid<sup>12</sup> (gift from N. Mizushima) with use of Lipofectamine LTX (Invitrogen) following the manufacturer's instructions. Accumulation of GFP-LC3 punctae or EYFP signal was assessed using LSM510 META confocal microscopy (Carl Zeiss).

**Intestinal epithelial cell purification and crypt isolation.** Mice were euthanized and the intestine was washed with ice-cold PBS after being cut open longitudinally. Peyer's patches were removed and the intestine was cut into small pieces. Mucus was removed by shaking the intestine in 1 $\times$  HBSS containing 1 mM DTT for 10 min at room temperature. After washing with PBS, pieces were digested with dispase (1 U ml<sup>-1</sup> in RPMI with 2% FCS) for 30 min at 37 °C with shaking (250 r.p.m.). Cells were collected and debris removed with a 100  $\mu$ m cell strainer, and centrifuged for 5 min at 1,500 r.p.m. Intestinal epithelial cells were collected in the top layer after 40–100% Percoll gradient centrifugation. Purity of the population was determined by staining with anti-EpCAM antibody and flow cytometry analysis. Intestinal epithelial cells were lysed with RIPA buffer and equal amounts of protein were used for western blot analysis as indicated in figure legends. To isolate small intestinal crypts, the intestine was flushed, cut open longitudinally and incubated on ice for 30 min in 2 mM EDTA/PBS. Two sedimentation steps and application of a cell strainer separated crypts from villi<sup>42</sup>, which were then used for RNA isolation (Qiagen), *Xbp1* splicing assay, and for protein lysis with RIPA buffer and subsequent immunoblotting.

**Intestinal epithelial scrapings.** Mice were euthanized, intestines collected and longitudinally opened, and immediately washed with ice-cold PBS. Intestinal epithelium was collected by scraping with glass slides and snap frozen into liquid nitrogen for further analysis. For protein analysis, intestinal epithelial scrapings were homogenized in RIPA buffer using a 25G needle with a syringe. Lysates were cleared by centrifugation and aliquots of protein were used for protein assessment using standard western blot or immunoprecipitation protocols as indicated. Isolation of mRNA from intestinal epithelial scrapings or MODE-K lysates and

RT-qPCR was performed as described<sup>4</sup> using following pairs of primers. *grp78* 5'-ACTTGGGGACCACCTATTCT-3' and 5'-ATCGCAATCAGACGCTCC-3'; *Gadd34* 5'-CCCGAGATTCTCTCTAAAAGC-3' and 5'-CCAGACAGCAAGGAAATGG-3'; *LC3b* 5'-GCGCCATGCCGTCGCGAGAAG-3' and 5'-GCTCCCGATGAGCCGGACA-3'; *Xbp1* 5'-AGCAGCAAGTGGTGGATTG-3' and 5'-GAGTTTTCTCCGTAAGAGCTGA-3'; *Atg7* 5'-CCTTCGCGGACCTAAAGAAGT-3' and 5'-CCCGATTAGAGGGATGCTC-3'; *Nfkb1a* 5'-TGAAGGACGACGAGTACGAGC-3' and 5'-TTCGTGGATGATTGCCAAGTG-3';  $\beta$ -actin 5'-GACGGCCAGGTCATCACTATTG-3' and 5'-AGGAAGGCTGGAAAAGA GCC-3'.

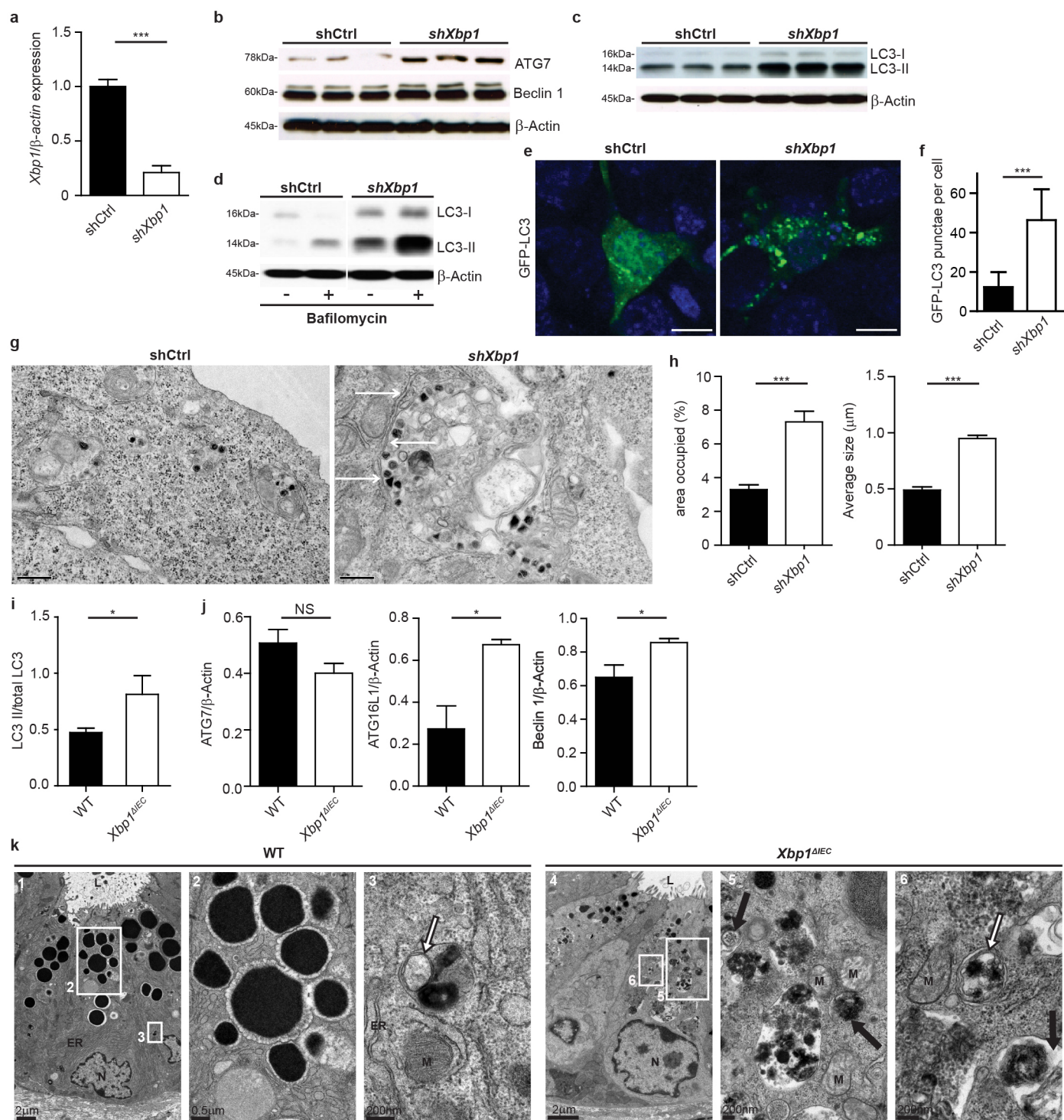
**Dextran sodium sulphate induced colitis.** Acute colitis was induced by adding 4% DSS (TDB Consultancy) to drinking water ad libitum for five consecutive days. Daily disease activity index (DAI) was assessed evaluating weight loss, stool consistency and rectal bleeding according to Supplementary Table 1. A high resolution mouse endoscopic system (Hopkins) was used. Colitis severity was assessed by a semi-quantitative score consisting of two subscores; endoscopic tissue damage (0–3, where 0 is no damage, 1 is lymphoepithelial lesions, 2 is surface mucosal erosion or focal ulceration, and 3 is extensive mucosal damage with expansion into deeper structures of the bowel wall) and inflammatory infiltration (0–3, where 0 is occasional inflammatory cells in the lamina propria, 1 is increased numbers on inflammatory cells in lamina propria, 2 is confluence of inflammatory cells extending into the submucosa, and 3 is transmural extension of the infiltrate).

***Xbp1* splicing assay and densitometric quantification.** *Xbp1* splicing was assessed as described previously<sup>4</sup>. Briefly, RNA was isolated, reverse transcribed and amplified by RT-PCR with the following primers: *Xbp1* sp. forward 5'-ACACGCTTGGGAATGGACAC-3'; *Xbp1* sp. reverse 5'-CCATGGGAAGATGTTCTGGG-3'. The PCR product of 171 (unspliced) and 145 (spliced) bp were resolved on a 2% agarose gel. Densitometric analysis for splicing assay and immunoblots was performed with ImageJ.

**Statistical methods.** Statistical significance was calculated as appropriate using an unpaired two-tailed Student's *t*-test or a Mann-Whitney *U*-test and considered significant at *P* < 0.05. In experiments where more than two groups were compared, Kruskal-Wallis test followed by Mann-Whitney *U*-test and post-hoc Bonferroni Holm's correction or one-way ANOVA/Bonferroni was performed. Grubb's test was used as appropriate to identify outliers. Data were analysed using GraphPad Prism software. Experimental group sizes were based on the goal of achieving desired effect sizes typically of  $\leq 2.0$  standard deviations and a power of 0.9 on the assumption of a normal distribution, and therefore typically involved *n* = 6–10.

- Madison, B. B. *et al.* Cis elements of the villin gene control expression in restricted domains of the vertical (crypt) and horizontal (duodenum, cecum) axes of the intestine. *J. Biol. Chem.* **277**, 33275–33283 (2002).
- Shimshek, D. R. *et al.* Codon-improved Cre recombinase (iCre) expression in the mouse. *Genesis* **32**, 19–26 (2002).
- Garabedian, E. M., Roberts, L. J., McNeven, M. S. & Gordon, J. I. Examining the role of Paneth cells in the small intestine by lineage ablation in transgenic mice. *J. Biol. Chem.* **272**, 23729–23740 (1997).
- Bry, L. *et al.* Paneth cell differentiation in the developing intestine of normal and transgenic mice. *Proc. Natl Acad. Sci. USA* **91**, 10335–10339 (1994).
- Srinivas, S. *et al.* Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev. Biol.* **1**, 4 (2001).
- Iwakaki, T., Akai, R., Yamanaka, S. & Kohno, K. Function of IRE1  $\alpha$  in the placenta is essential for placental development and embryonic viability. *Proc. Natl Acad. Sci. USA* **106**, 16657–16662 (2009).
- Baert, L. *et al.* Detection of murine norovirus 1 by using plaque assay, transfection assay, and real-time reverse transcription-PCR before and after heat exposure. *Appl. Environ. Microbiol.* **74**, 543–546 (2008).
- Pierce, J. W. *et al.* Novel inhibitors of cytokine-induced I $\kappa$ B $\alpha$  phosphorylation and endothelial cell adhesion molecule expression show anti-inflammatory effects *in vivo*. *J. Biol. Chem.* **272**, 21096–21103 (1997).
- Satoh, Y., Yamano, M., Matsuda, M. & Ono, K. Ultrastructure of Paneth cells in the intestine of various mammals. *J. Electron Microsc. Tech.* **16**, 69–80 (1990).
- Vidal, K., Grosjean, I., Revillard, J. P., Gespach, C. & Kaiserlian, D. Immortalization of mouse intestinal epithelial cells by the SV40-large T gene. Phenotypic and immune characterization of the MODE-K cell line. *J. Immunol. Methods* **166**, 63–73 (1993).
- Hempel, S. L., Buettner, G. R., O'Malley, Y. Q., Wessels, D. A. & Flaherty, D. M. Dihydrofluorescein diacetate is superior for detecting intracellular oxidants: comparison with 2',7'-dichlorodihydrofluorescein diacetate, 5-(and 6)-carboxy-2',7'-dichlorodihydrofluorescein diacetate, and dihydrorhodamine 123. *Free Radic. Biol. Med.* **27**, 146–159 (1999).
- Sato, T. *et al.* Single Lgr5 stem cells build crypt-villus structures *in vitro* without a mesenchymal niche. *Nature* **459**, 262–265 (2009).
- Wei, Y., Sinha, S. & Levine, B. Dual role of JNK1-mediated phosphorylation of Bcl-2 in autophagy and apoptosis regulation. *Autophagy* **4**, 949–951 (2008).
- Carchman, E. H., Rao, J., Loughran, P. A., Rosengart, M. R. & Zuckerbraun, B. S. Heme oxygenase-1-mediated autophagy protects against hepatocyte cell death and hepatic injury from infection/sepsis in mice. *Hepatology* **53**, 2053–2062 (2011).

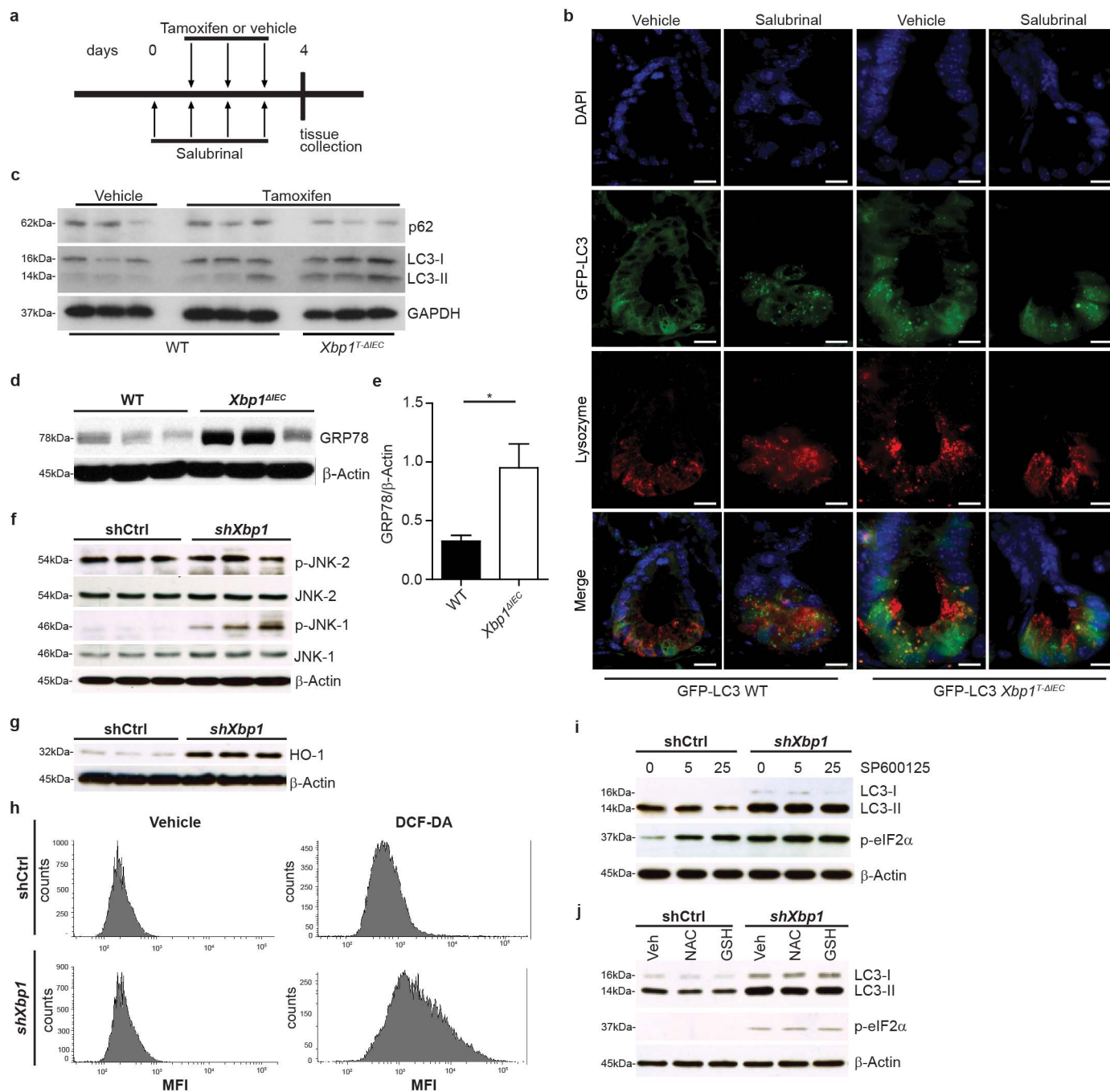
45. Hu, P., Han, Z., Couvillon, A. D., Kaufman, R. J. & Exton, J. H. Autocrine tumor necrosis factor alpha links endoplasmic reticulum stress to the membrane death receptor pathway through IRE1 $\alpha$ -mediated NF- $\kappa$ B activation and down-regulation of TRAF2 expression. *Mol. Cell. Biol.* **26**, 3071–3084 (2006).
46. Urano, F. *et al.* Coupling of stress in the ER to activation of JNK protein kinases by transmembrane protein kinase IRE1. *Science* **287**, 664–666 (2000).
47. Barrett, J. C. *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nature Genet.* **40**, 955–962 (2008).
48. Tashiro, E. *et al.* Trierixin, a novel Inhibitor of ER stress-induced XBP1 activation from *Streptomyces* sp. 1. Taxonomy, fermentation, isolation and biological activities. *J. Antibiot.* **60**, 547–553 (2007).
49. Rioux, J. D. *et al.* Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nature Genet.* **39**, 596–604 (2007).
50. Parkes, M. *et al.* Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nature Genet.* **39**, 830–832 (2007).





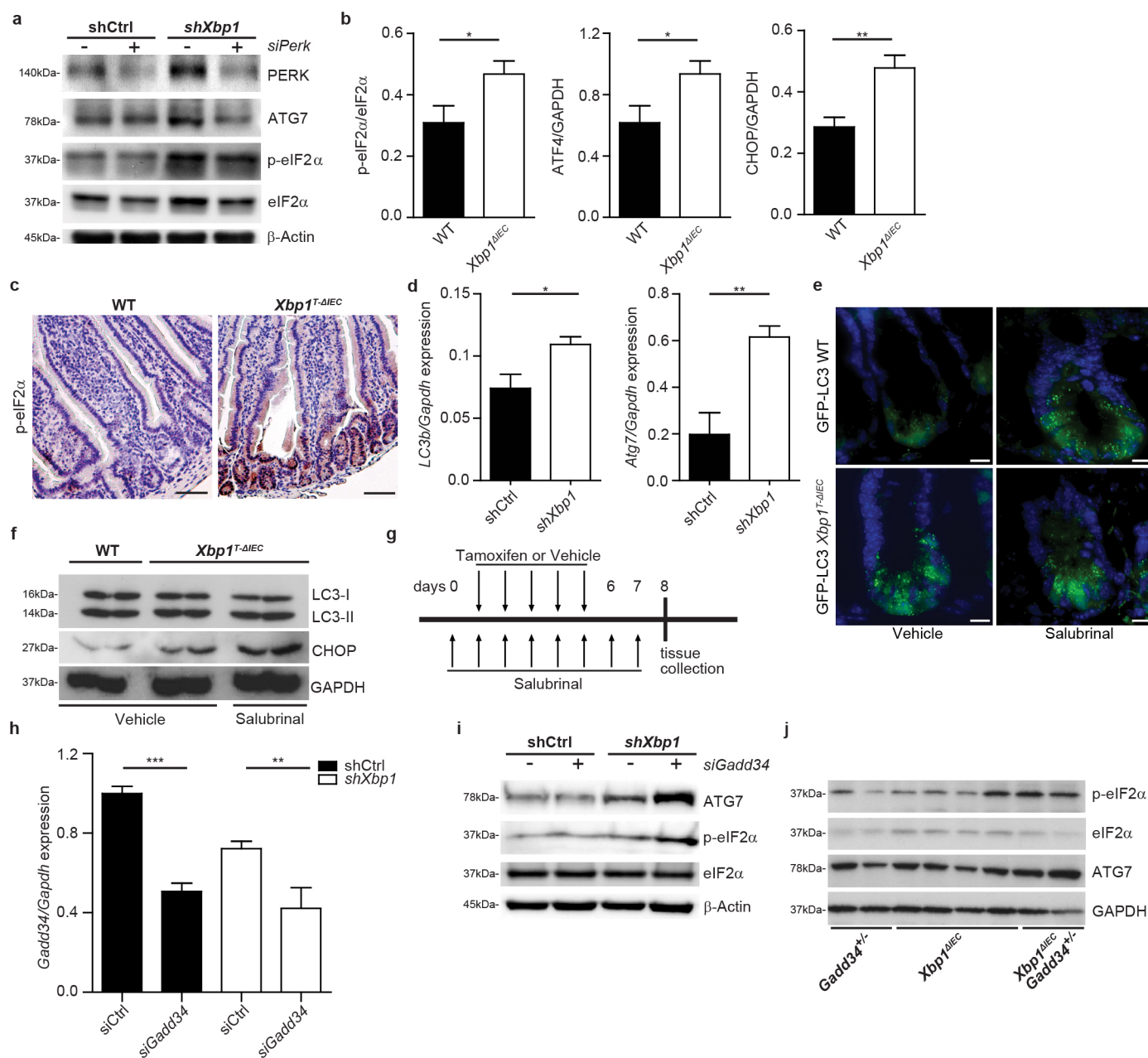
**Extended Data Figure 1 | Autophagy induction in *Xbp1*-deficient intestinal epithelial cells.** **a**, *Xbp1* expression in shCtrl and *shXbp1* MODE-K cells ( $n = 7/6$ ; unpaired Student's *t*-test; mean  $\pm$  s.e.m.). **b, c**, Immunoblot of shCtrl and *shXbp1* MODE-K cells. **d**, Immunoblot after autophagosome-lysosome fusion inhibition via bafilomycin in silenced MODE-K intestinal epithelial cells. **e**, Silenced MODE-K intestinal epithelial cells after GFP-LC3 reporter transfection (scale bars, 5  $\mu$ m) with green punctae per cell quantification in **f** ( $n = 14$ ; unpaired Student's *t*-test; mean  $\pm$  s.e.m.). **g**, TEM of shCtrl and *shXbp1* cells. Note double-membraned structure with engulfed contents characteristic of autophagosomes (white arrows;  $n = 10$ ). Scale bars, 0.5  $\mu$ m. **h**, Quantification of occupied area and average size of autophagic vacuoles from **g** ( $n = 10$ ; unpaired Student's *t*-test; mean  $\pm$  s.e.m.). **i, j**, Densitometry of Fig. 1a (**i**;  $n = 5/4$ ; unpaired Student's *t*-test; mean  $\pm$  s.e.m.) and Fig. 1b (**j**;  $n = 3$ ; unpaired Student's *t*-test; mean  $\pm$  s.e.m.). **k**, Low-magnification ( $\times 1,380$ , original magnification here and in the remainder of this legend) TEM image of Paneth cells from wild-type mice (1), demonstrating the abundant endoplasmic

reticulum (ER) and characteristic secretory granules at the apical, lumenally (L) oriented side. (2) Higher magnification ( $\times 5,520$ ) of inset '2' in (1) demonstrating typical secretory granules in Paneth cells from wild-type mice. (3) High magnification ( $\times 20,700$ ) of inset '3' in (1) illustrating a double membrane structure characteristic of an autophagosome (white arrow) in close proximity to the ER and a mitochondrion. (4) Low magnification ( $\times 2,160$ ) TEM image of Paneth cell remnants present in *Xbp1*<sup>ΔIEC</sup> small intestinal crypts, which lack expansion of the ER and exhibit only minuscule granule remnants. (5) Higher magnification ( $\times 9,000$ ) of inset '5' in (4), demonstrating degradative autophagic vacuoles (black arrows), in close proximity to mitochondria, and the virtual absence of ER membranes. (6) High-power ( $\times 14,400$ ) magnification of inset '6' in (4), illustrating a double-membrane structure (white arrow) characteristic of autophagosomes, and a degradative autophagic vacuole (black arrow). ER, endoplasmic reticulum; L, lumen; M, mitochondrion; N, nucleus. Scale bars represent 2  $\mu$ m, 0.5  $\mu$ m and 200 nm, respectively. Results represent three (**b, c**) or two (**d, k**) independent experiments. \* $P < 0.05$ , \*\*\* $P < 0.001$ .



**Extended Data Figure 2 | Autophagy in *Xbp1*-deficient intestinal epithelial cells is induced by eIF2 $\alpha$  phosphorylation but independent of JNK or oxidative stress pathways.** **a**, Timeline for salubrinal experiment shown in Fig. 1j. **b**, Representative indirect immunofluorescence images of GFP-LC3 puncta accumulation (green) in small intestinal sections co-stained with anti-lysosome antibody (red) treated as in **a** ( $n = 3$ ). DAPI, blue. Scale bars, 10  $\mu$ m. **c**, Immunoblot of crypt lysates after 3 days of tamoxifen or vehicle administration ( $n = 3$ ). **d**, **e**, Immunoblot of primary intestinal epithelial cell scrapings (**d**); densitometry in **e** ( $n = 3$ ; unpaired Student's *t*-test; mean  $\pm$  s.e.m.). **f**, **g**, Immunoblot of *shXbp1* and *shCtrl* MODE-K cells. ER stress-induced Jun N-terminal kinase-1 (JNK1) has previously been connected in other cellular model systems to autophagy activation through phosphorylation of B-cell leukaemia 2 (Bcl-2) and its dissociation from beclin

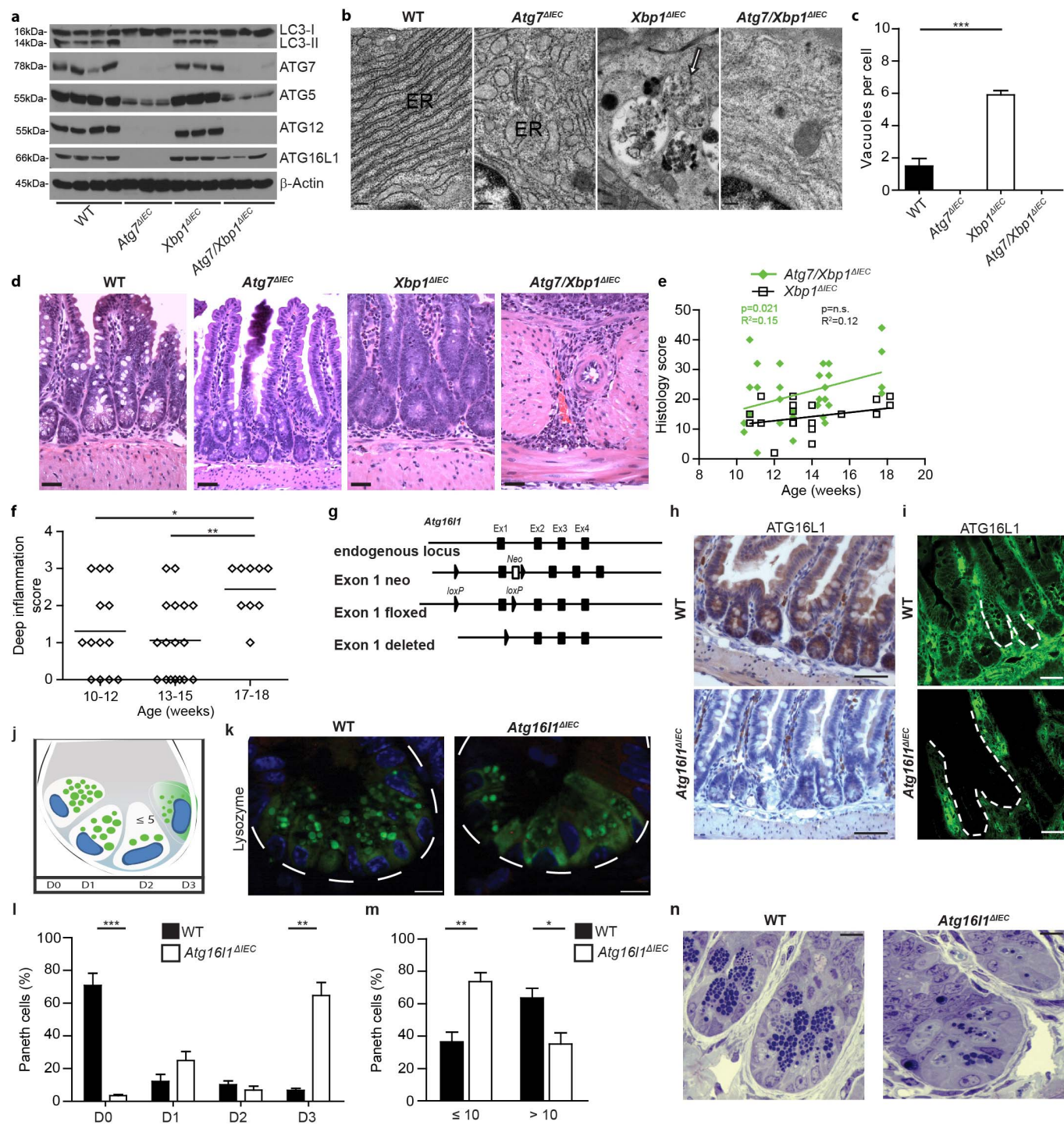
<sup>143</sup>, as have oxidative stress/free radicals and haem oxygenase-1 (HO-1) activation<sup>44</sup>. **h**, Intracellular ROS determined by dichlorofluorescein assay and mean fluorescent intensity (MFI) after vehicle or dichlorofluorescein diacetate (DCF-DA) treatment. **i**, Immunoblot of *shXbp1* and *shCtrl* MODE-K cells after administration of the JNK inhibitor SP600125 (0, 5 or 25  $\mu$ M) for 4 h. Note the absence of an effect of SP600125 treatment on the conversion of LC3-I to LC3-II or the levels of p-eIF2 $\alpha$ , thereby excluding a major contribution of the JNK pathway to autophagy induction in the presence of intestinal-epithelial-cell-associated XBP1 deficiency. **j**, Immunoblot of *shXbp1* and *shCtrl* MODE-K cells after *N*-acetylcysteine (NAC), glutathione (GSH) or vehicle treatment for 16 h. Note the absence of an effect of the free radical scavengers on either of these markers of UPR-induced autophagy (LC3-II or p-eIF2 $\alpha$ ). Results represent three (f–j) independent experiments. \* $P < 0.05$ .



**Extended Data Figure 3 | ER stress-induced activation of PERK/eIF2 $\alpha$  induces autophagy in *Xbp1*-deficiency.** **a**, Immunoblot of *shXbp1* and *shCtrl* MODE-K cells co-silenced with *Perk* (*siPerk*) or scrambled siRNA. **b**, Cumulative densitometry of the immunoblot in Fig. 1g and two additional experiments (not shown,  $n = 10$ ; unpaired Student's  $t$ -test; mean  $\pm$  s.e.m.). **c**, p-eIF2 $\alpha$  immunohistochemistry of small intestinal epithelium. Scale bars, 50  $\mu$ m. **d**, mRNA expression levels of *Map1lc3b* (*LC3b*) and *Atg7* relative to *Gapdh* in *shCtrl* and *shXbp1* MODE-K cells ( $n = 3$ ; unpaired Student's  $t$ -test; mean  $\pm$  s.e.m.). **e**, Accumulation of GFP-LC3 punctae after salubrinal and 3-day tamoxifen treatment according to timeline in Extended Data Fig. 2a

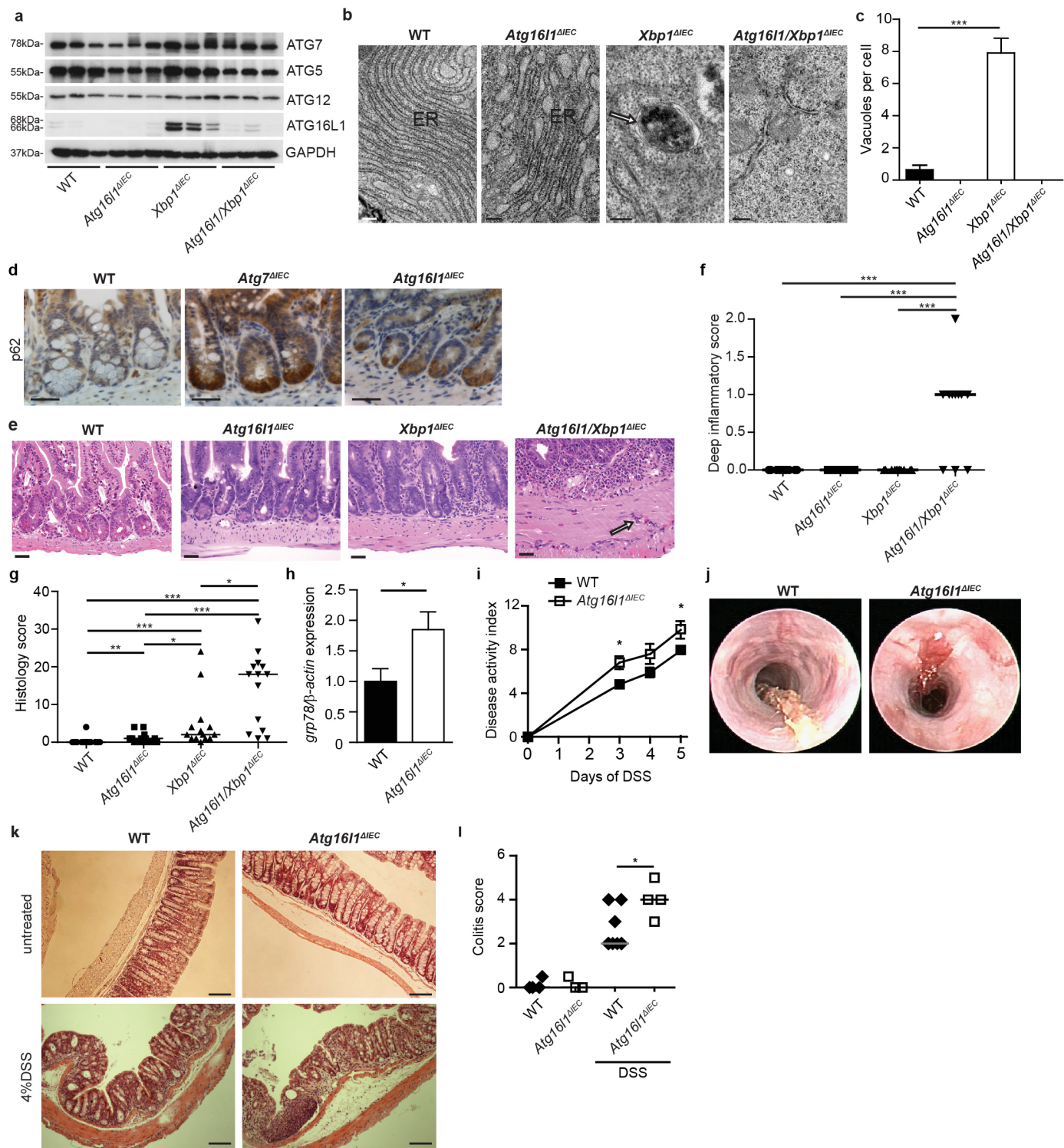
( $n = 5$ ). Scale bars, 5  $\mu$ m. **f**, Immunoblot of primary epithelial cell scrapings from small intestine upon vehicle or salubrinal treatment according to the timeline in **g** ( $n = 2$ ). Note the increased relative levels of LC3 conversion and CHOP, a transcriptional target of ATF4, in salubrinal-treated mice. **g**, Timeline of salubrinal experiment. Results are reported in **f** and Fig. 1k. **h**, *Gadd34* expression in *shCtrl* and *shXbp1* MODE-K cells co-silenced for *Gadd34* (*siGadd34*) and control (*siCtrl*) (one-way ANOVA with post-hoc Holm's corrected unpaired Student's  $t$ -test; mean  $\pm$  s.e.m.). **i**, Cells as in **h** were immunoblotted. **j**, Immunoblot of crypt lysates ( $n = 2/4/2$ ). Results represent three (**a**, **h**, **i**) independent experiments. \* $P$  < 0.05, \*\* $P$  < 0.01, \*\*\* $P$  < 0.001.





**Extended Data Figure 4 | Absence of *Atg7* in intestinal epithelium disrupts autophagy and exacerbates spontaneous enteritis of *Xbp1* deficiency and development of conditional *Atg16l1* knockout mouse.** **a**, Immunoblot of primary intestinal epithelial cell scrapings ( $n = 4/3/3/3$ ). Note the loss of ATG5, ATG12 and ATG16L1 proteins in ATG7-deficient mice. **b**, Representative high magnification TEM images ( $n = 2$ ). Note intact ER in wild-type mice; severe distortion of ER in *Atg7*<sup>AIEC</sup> mice; autophagic vacuoles with the characteristic double membrane (white arrow) in *Xbp1*<sup>AIEC</sup> mice; and the absence of ER and autophagic vacuoles in *Atg7/Xbp1*<sup>AIEC</sup> mice. Scale bars, 200 nm. ER, endoplasmic reticulum. **c**, Quantification of autophagic vacuoles ( $n = 10$ ; unpaired Student's *t*-test; mean  $\pm$  s.e.m.). **d**, High-magnification ( $\times 400$ ) haematoxylin and eosin images. Note knife-like extension of inflammation along blood vessels through whole thickness of muscularis propria into the serosa in *Atg7/Xbp1*<sup>AIEC</sup> mice. Scale bars, 10  $\mu$ m. **e**, Linear regression analysis for the correlation of inflammation with age. Each dot represents a single animal ( $R^2$  and *P* value for deviation from 0 are shown). **f**, Deep inflammation score (as described in Methods) in *Atg7/Xbp1*<sup>AIEC</sup> mice plotted at indicated ages in weeks ( $n = 13/17/9$ ; median shown; Kruskal–Wallis with post-hoc Holm's-corrected Mann–Whitney *U*-test). **g**, *Atg16l1* intestinal-epithelial-cell-specific knockout design. Exon 1 of the *Atg16l1* gene was flanked by two *loxP* sites and *Atg16l1* deletion was mediated by Cre recombinase (Cre) under

control of the Villin (V-) promoter (*V-cre*<sup>+</sup>; *Atg16l1*<sup>H/J</sup> or *Atg16l1*<sup>AIEC</sup>). See Methods for a detailed description. **h**, Immunohistochemical staining (brown) with anti-ATG16L1 antibody (MBL) ( $n = 2$ ). Note cytoplasmic intestinal-epithelial-cell-specific ATG16L1 immunostaining in wild-type but not *Atg16l1*<sup>AIEC</sup> mice with retained ATG16L1 expression in lamina propria mononuclear cells in *Atg16l1*<sup>AIEC</sup> mice. Scale bars, 50  $\mu$ m. **i**, Indirect immunofluorescence microscopy (green) with anti-ATG16L1 antibody (Cell Signaling Technology) ( $n = 2$ ). Note the similar staining pattern as in **h** with cytoplasmic intestinal-epithelial-cell-specific ATG16L1 immunoreactivity in wild-type but not *Atg16l1*<sup>AIEC</sup> mice and retained ATG16L1 expression in lamina propria mononuclear cells in *Atg16l1*<sup>AIEC</sup> mice. Dashed line denotes crypt unit. Scale bars, 20  $\mu$ m. **j**, Schematic representation of lysozyme<sup>+</sup> granule allocation patterns (green): normal (D0), disordered (D1), depleted (D2), and diffuse (D3)<sup>2</sup>. **k, l**, Lysozyme immunofluorescence (green) in crypts of *Atg16l1*<sup>AIEC</sup> and wild-type mice (**k**), quantified in **l** ( $n = 3$ ; unpaired Student's *t*-test; mean  $\pm$  s.e.m.) according to granule allocation patterns shown in **j**. Dashed line denotes crypt unit. DAPI, blue. Scale bars, 5  $\mu$ m. **m, n**, Quantification of Paneth cells based on the number of granular vesicles stained with toluidine blue (**m**) as shown in **n** ( $n = 4$ ; unpaired Student's *t*-test; mean  $\pm$  s.e.m.). Scale bars, 5  $\mu$ m. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

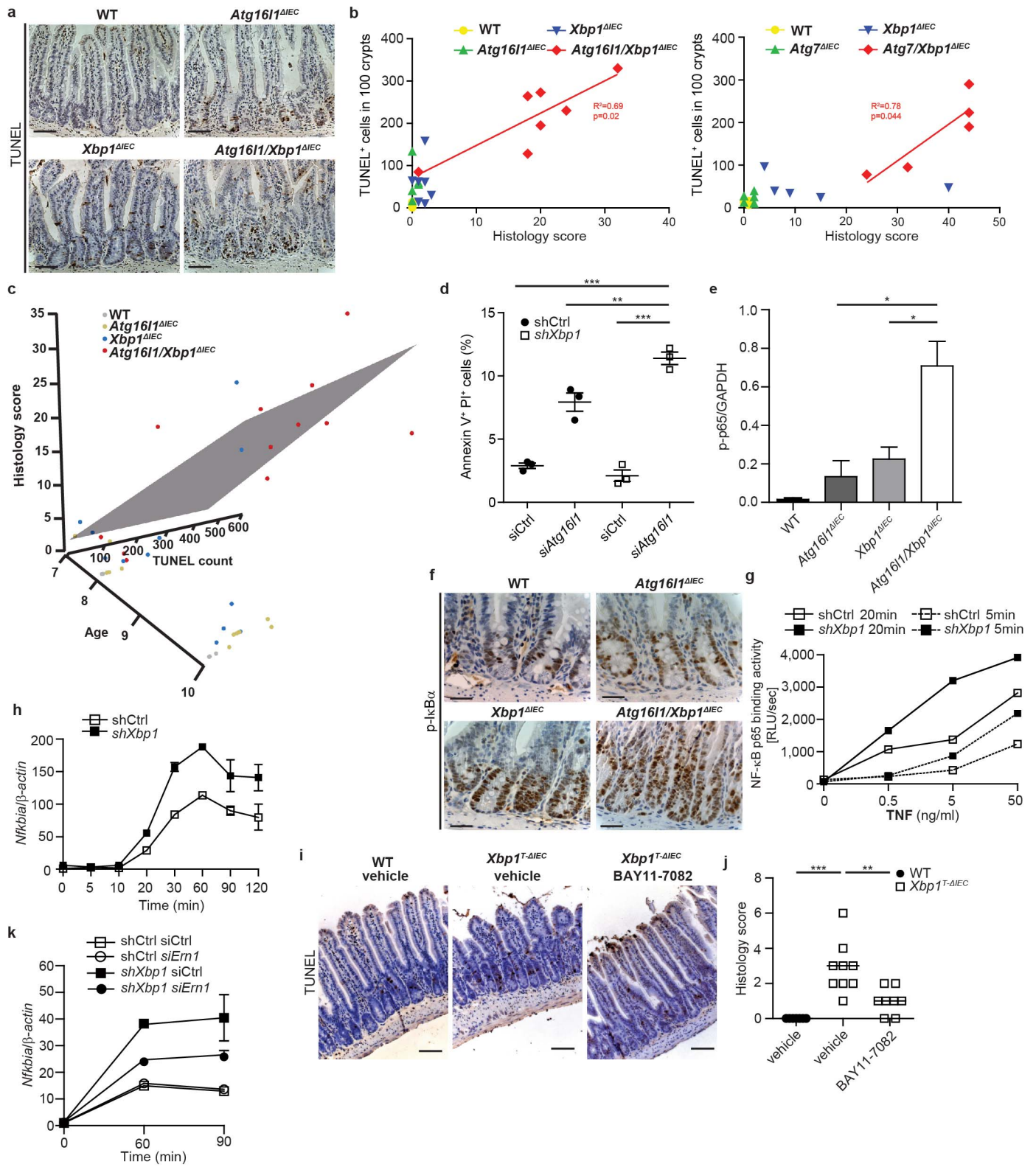




**Extended Data Figure 5 | *Atg16l1* intestinal-epithelial-cell-specific deletion abrogates UPR-induced autophagy and exacerbates ER stress-induced spontaneous intestinal inflammation due to *Xbp1* deficiency.**

**a**, Immunoblot of primary intestinal epithelial cell scrapings. Note identical GAPDH loading control as in Fig. 3b ( $n = 3$ ). **b**, Representative high-magnification TEM images ( $n = 2$ ). Note intact ER in wild-type mice; severe distortion of ER in *Atg16l1*<sup>AIEC</sup> mice; autophagic vacuoles with the characteristic double membrane (arrow) in *Xbp1*<sup>AIEC</sup> mice; and absence of ER and autophagic vacuoles in *Atg16l1/Xbp1*<sup>AIEC</sup> mice. Scale bar, 200 nm. ER, endoplasmic reticulum. **c**, Quantification of autophagic vacuoles ( $n = 10$  unpaired Student's *t*-test; mean  $\pm$  s.e.m.). **d**, Representative images of p62 immunostaining on small intestinal sections ( $n = 3$ ). Brown staining indicates p62 with maximal signal intensity in the region of Paneth cells. Scale bars, 20  $\mu$ m. **e**, High-resolution haematoxylin and eosin images. Note the extension

of the inflammation into muscularis propria (arrow) in *Atg16l1/Xbp1*<sup>AIEC</sup> mice. Scale bars, 20  $\mu$ m. **f**, Deep inflammation score (as described in Methods) in mice assessed at 18 weeks of age ( $n = 11$ ; median shown; Kruskal–Wallis with post-hoc Dunn's correction). **g**, Enteritis histology score ( $n = 14$ ; 7–10-week-old mice housed in Innsbruck; median shown; Kruskal–Wallis with post-hoc Holm's-corrected Mann–Whitney *U*-test). **h**, qRT–PCR of *grp78* in primary intestinal epithelial cell scrapings ( $n = 7$ ; unpaired Student's *t*-test; mean  $\pm$  s.e.m.). **i**, Disease activity index during DSS colitis ( $n = 7/4$ , two-way ANOVA with post-hoc Bonferroni; mean  $\pm$  s.e.m.). **j**, Representative endoscopic images from the colon at day 5 of DSS treatment ( $n = 3$ ). **k**, Representative haematoxylin and eosin staining of colonic sections from 5-day DSS-treated and untreated mice ( $n = 7/4$ ). Scale bars, 50  $\mu$ m. **l**, Colitis score after 5 days of DSS administration ( $n = 7/4$ , Mann–Whitney *U*-test). \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

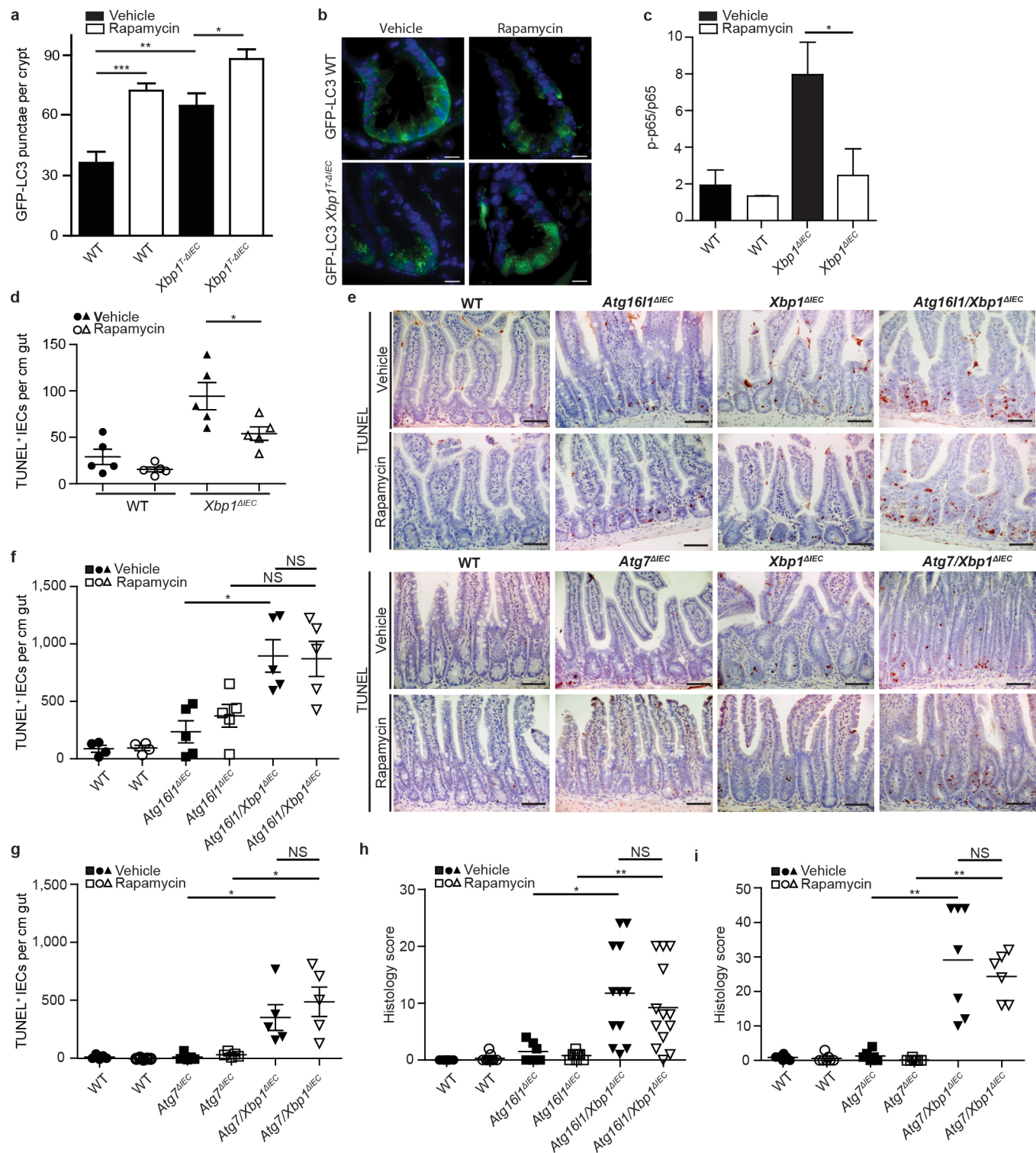


**Extended Data Figure 6 | ER stress-induced enteritis of *Xbp1* deficiency exacerbated by autophagy impairment correlates with age and cell death in intestinal epithelial cells and is dependent on IRE1 $\alpha$ -mediated NF- $\kappa$ B activation.**

**a**, Representative images of TUNEL-labelled intestinal epithelium (brown) ( $n = 7$ ). Scale bars, 50  $\mu$ m. **b**, Correlation of TUNEL<sup>+</sup> cells with the severity of inflammation. Linear regression analysis of *Atg16l1/Xbp1*<sup>AIEC</sup> (left) and *Atg7/Xbp1*<sup>AIEC</sup> (right) with significant  $R^2$  and  $P$  value for deviation from zero shown (left panel,  $n = 5/5/7/7$ ; right panel,  $n = 5$ ). **c**, Linear least square regression analysis for the correlation of enteritis histology score with cell death and age of animals by genotype. Each dot represents a single animal (grey, wild type; yellow, *Atg16l1*<sup>AIEC</sup>; blue, *Xbp1*<sup>AIEC</sup>; red, *Atg16l1/Xbp1*<sup>AIEC</sup> mice) and the plane represents the linear regression for the enteritis histology score as a function of age and TUNEL labelling for *Atg16l1/Xbp1*<sup>AIEC</sup> mice ( $n = 6/13/12/12$ ). Note that the severity of inflammation significantly correlates with numbers of TUNEL<sup>+</sup> intestinal epithelial cells and age only in *Atg16l1/Xbp1*<sup>AIEC</sup> mice ( $R^2 = 0.602$ ,  $P = 0.016$ ). The three-dimensional (3D) plot is also available online in video format (Supplementary Video 1). Regression analysis was performed using the R package lessR (<http://cran.r-project.org/web/packages/lessR/index.html>), last accessed May 2013. **d**, shCtrl or *shXbp1*

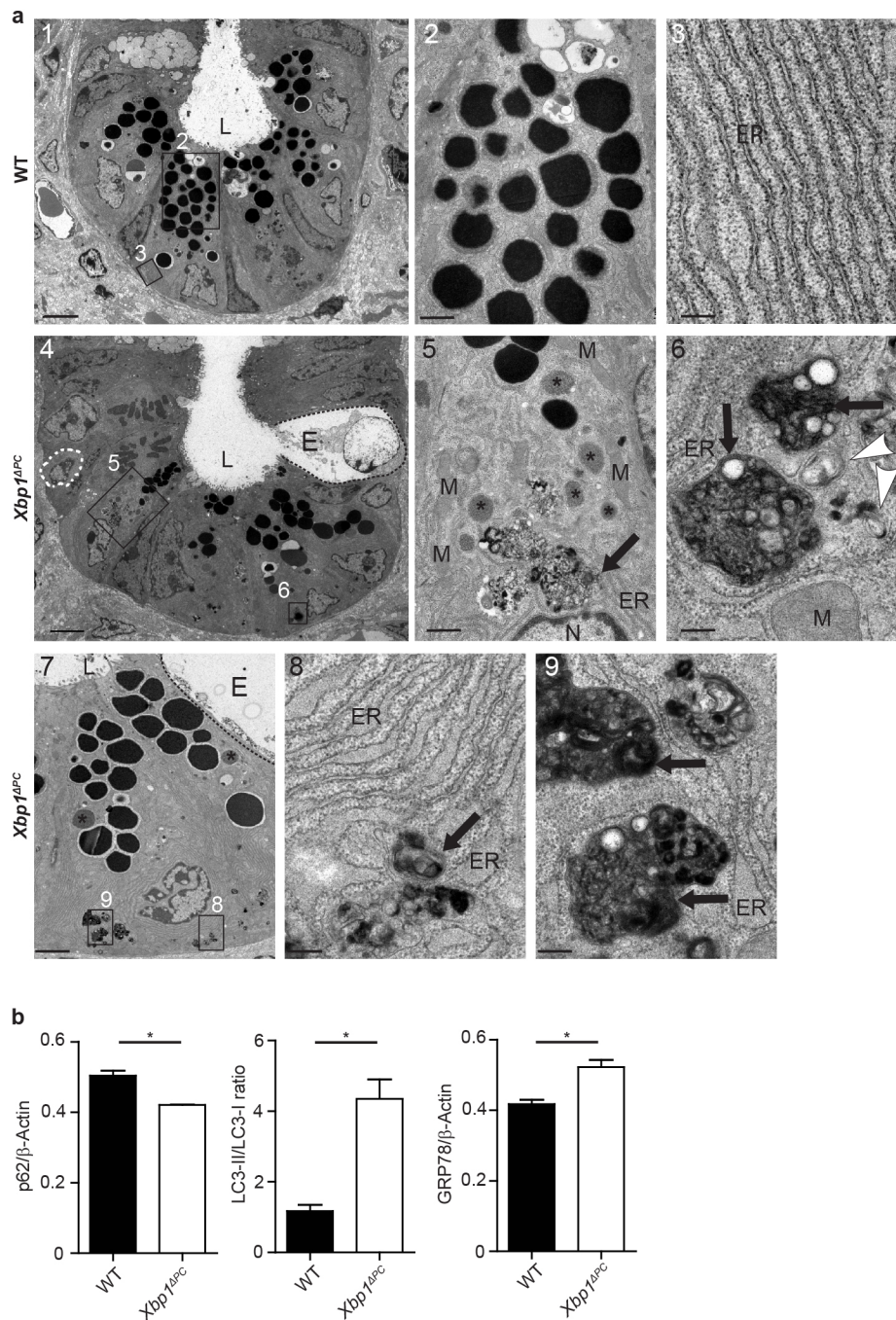
MODE-K cells were co-silenced for *Atg16l1* (*siAtg16l1*) or with scrambled siRNA (siCtrl), and analysed by flow cytometry for annexin V and propidium iodide (PI) uptake (one-way ANOVA with post-hoc Bonferroni; mean  $\pm$  s.e.m.). **e**, Densitometry of the immunoblot shown in Fig. 3b ( $n = 3$ ; one-way ANOVA with post-hoc Holm's corrected unpaired Student's  $t$ -test; mean  $\pm$  s.e.m.). **f**, Immunohistochemical staining for p-I $\kappa$ B $\alpha$  in the small intestinal epithelium ( $n = 3$ ). Scale bars, 20  $\mu$ m. **g**, NF- $\kappa$ B consensus sequence binding assay after stimulation of shCtrl and *shXbp1* MODE-K cells for 5 and 20 min with indicated concentrations of TNF. **h**, *Nfkb1a* expression, a prototypic NF- $\kappa$ B-transactivated gene, after TNF stimulation of shCtrl and *shXbp1* MODE-K cells (mean  $\pm$  s.e.m.). **i**, Representative images of TUNEL-labelled sections after administration of BAY11-7082 or vehicle ( $n = 3/4/4$ ). Scale bars, 50  $\mu$ m. **j**, Enteritis histology score of mice treated with the NF- $\kappa$ B inhibitor BAY11-7082 or vehicle ( $n = 7/9/8$ ; median shown; Kruskal–Wallis with post-hoc Holm's-corrected Mann–Whitney  $U$ -test). **k**, Expression of *Nfkb1a* in *Ern1*- and control-silenced *shXbp1* and control MODE-K cells after TNF stimulation (mean  $\pm$  s.e.m.). Results represent three (**h**, **k**) or two (**d**, **g**) independent experiments. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .





**Extended Data Figure 7 | Rapamycin induces autophagy in Paneth cells, reduces the severity of enteritis and the number of TUNEL<sup>+</sup> intestinal epithelial cells in *Xbp1<sup>ΔIEC</sup>* but not in *Atg7/Xbp1<sup>ΔIEC</sup>* or *Atg161/Xbp1<sup>ΔIEC</sup>* mice.** **a, b**, Quantification of GFP-LC3 punctae accumulation after 3-day tamoxifen treatment and induction by rapamycin (**a**;  $n = 10$ ; one-way ANOVA with post-hoc Bonferroni; mean  $\pm$  s.e.m.) and representative images in **b** ( $n = 3$ ). Scale bars, 5  $\mu$ m. **c**, Densitometry of the immunoblot in Fig. 4a ( $n = 3$ ; one-way ANOVA with post-hoc Holm's-corrected unpaired Student's  $t$ -test; mean  $\pm$  s.e.m.). **d**, Quantification of TUNEL<sup>+</sup> intestinal epithelial cells per cm of gut in mice treated with rapamycin or vehicle ( $n = 5$ ; unpaired Student's

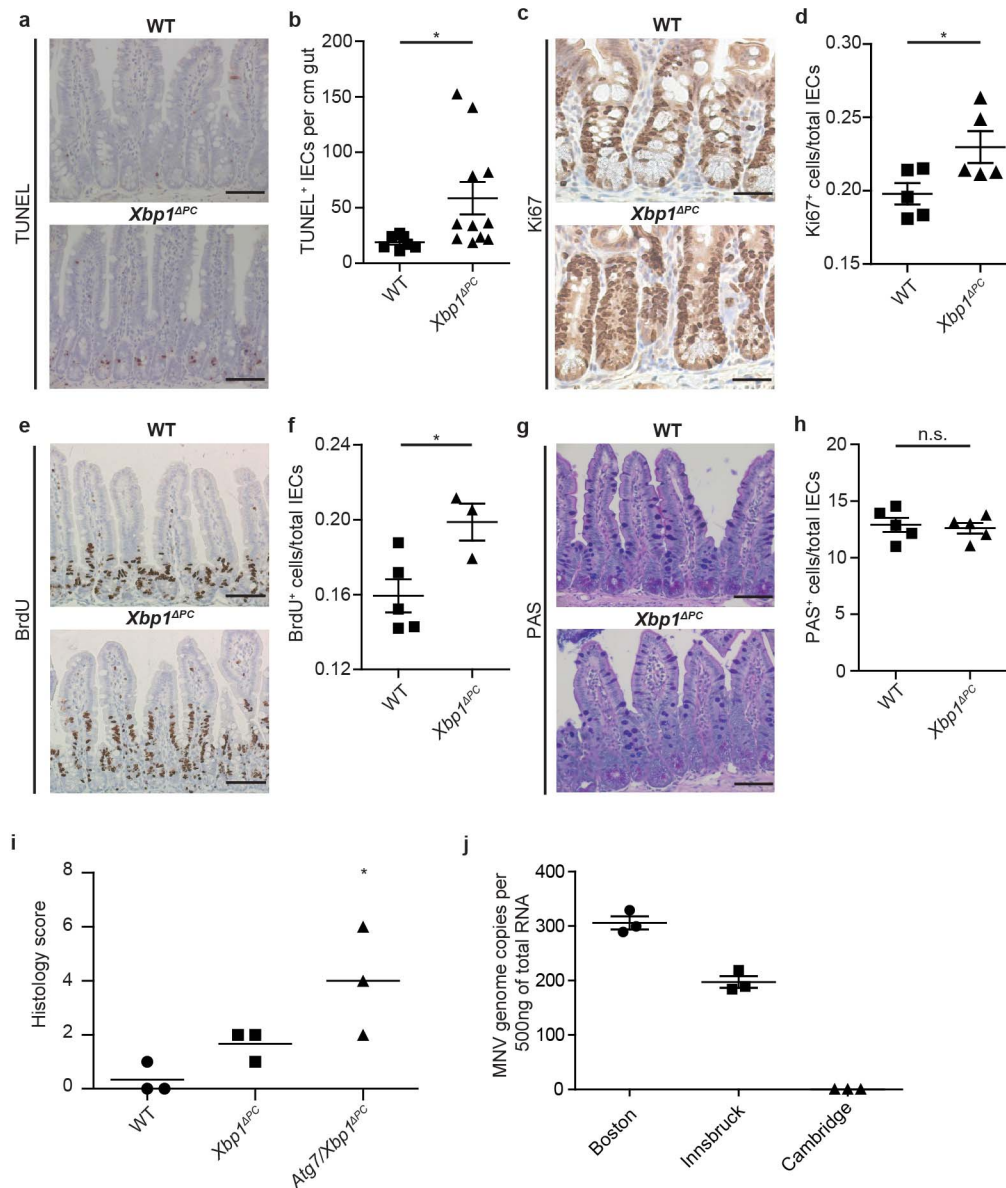
$t$ -test; mean  $\pm$  s.e.m.). **e**, Representative images of TUNEL-labelled sections from rapamycin- or vehicle-treated mice. Scale bars, 50  $\mu$ m. **f, g**, Quantification of TUNEL<sup>+</sup> intestinal epithelial cells per cm of gut in mice treated with rapamycin or vehicle (**f**;  $n = 4/4/5/5/5/5$ ; **g**;  $n = 5$ ; one-way ANOVA with post-hoc Holm's-corrected unpaired Student's  $t$ -test; mean  $\pm$  s.e.m.). **h, i**, Enteritis histology scores of rapamycin- or vehicle-treated mice of respective genotypes (**h**;  $n = 8/9/6/9/12/13$ ; **i**;  $n = 8/7/6/7/7/6$ ) (median shown; Kruskal-Wallis with post-hoc Holm's-corrected Mann-Whitney  $U$ -test). \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .



**Extended Data Figure 8 | Autophagy induction in Paneth cells of *Xbp1<sup>APC</sup>* mice.** **a**, Representative TEM images of small intestinal crypts from *Defa6-cre<sup>+</sup>* (wild type) and *Defa6-cre<sup>+</sup>;Xbp1<sup>fl/fl</sup>* (*Xbp1<sup>APC</sup>*) mice. (1) Low magnification TEM image of Paneth cells at the base of a small intestinal crypt in wild-type mice, demonstrating abundant ER and characteristic secretory granules at the apical, lumenally oriented side. Scale bar, 5  $\mu$ m. (2) Higher magnification of inset '2' in (1) with numerous secretory granules. Scale bar, 1  $\mu$ m. (3) High magnification of inset '3' in (1) with intact ER. Scale bar, 200 nm. (4) Low-magnification TEM image of Paneth cells in an *Xbp1<sup>APC</sup>* small intestinal crypt. Note the reduced number of secretory granules, the exfoliating hypomorphic Paneth cell into crypt lumen ('E', black dashed outline) and a transmigrating polymorphonuclear cell through the mucosa (white dashed outline). Scale bar, 5  $\mu$ m. (5) Higher magnification of inset '5' in (4), demonstrating abnormalities in secretory granule maturation with numerous hypodense granules (\*) and a distorted ER. Note the accumulation of electron-dense cargo indicative of degradative autophagic vacuoles

(black arrow) in close proximity to the nucleus. Scale bar, 1  $\mu$ m. (6) High magnification of the inset '6' in (4), demonstrating degradative autophagic vacuoles filled with electron-dense material (black arrow). Note the double membrane structures characteristic for autophagosomes (white arrowheads). Scale bar, 200 nm. (7) Low-magnification TEM image of Paneth cells in an *Xbp1<sup>APC</sup>* small intestinal crypt. Note the disorganized ER, hypodense secretory granules (\*), empty area after exfoliation of a Paneth cell into the crypt lumen ('E', black dashed outline) similar to image (4). (8) High magnification of inset '8' identified in (7), demonstrating an autophagic vacuole (black arrow) surrounded by a markedly distorted ER. Scale bar, 200 nm. (9) High magnification of the inset '9' in (7), with additional examples of degradative autophagic vacuoles accompanied by a distorted ER. Representative of two independent experiments. Scale bar, 200 nm. ER, endoplasmic reticulum; L, lumen; M, mitochondrion; N, nucleus. **b**, Densitometry of the immunoblot in Fig. 4d ( $n = 2$ ; unpaired Student's *t*-test; mean  $\pm$  s.e.m.). \* $P < 0.05$ .

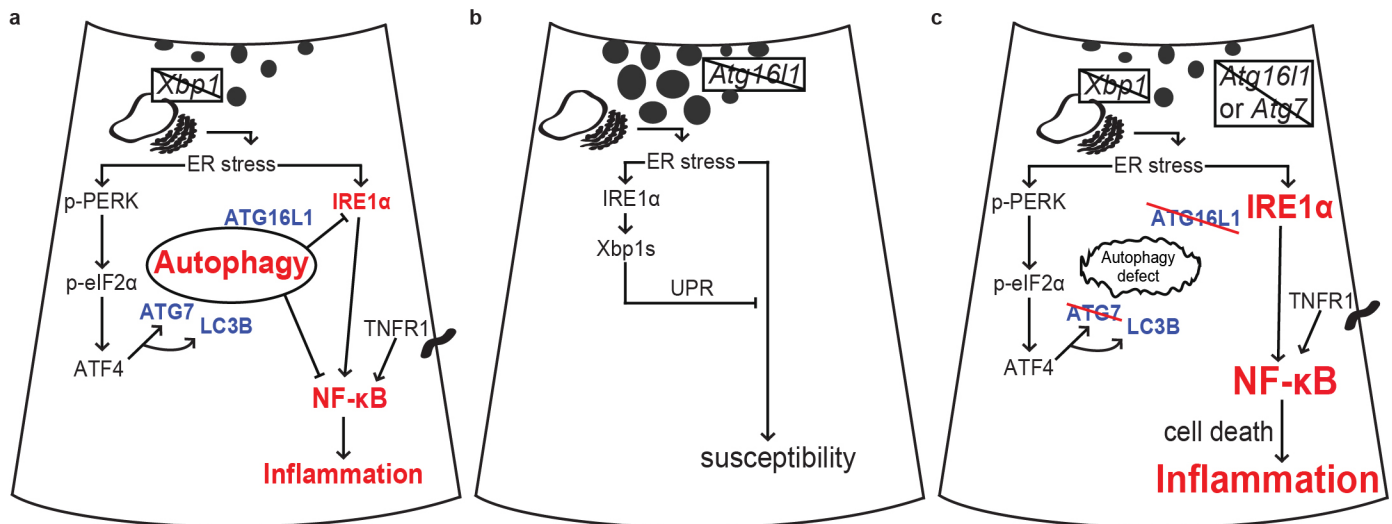




**Extended Data Figure 9 | Phenotypic consequences of *Xbp1* deletion in Paneth cells and characterization of murine norovirus status.** **a**, **b**, TUNEL-labelled ileal sections (**a**) with quantification of TUNEL<sup>+</sup> cells per cm gut shown in **b** ( $n = 7/11$ ; unpaired Student's *t*-test; mean ± s.e.m.). Scale bars, 50 μm. **c**, **d**, Ki67 immunohistochemical staining of ileal sections (**c**) with quantification of Ki67<sup>+</sup> cells per total intestinal epithelial cell count along the crypt–villus axis shown in **d** ( $n = 5$ ; unpaired Student's *t*-test; mean ± s.e.m.). Scale bars, 20 μm. **e**, **f**, BrdU-labelled ileal sections (**e**) after 24 h with quantification of BrdU<sup>+</sup> cells per total intestinal epithelial cell count along the crypt–villus axis in **f** ( $n = 5/3$ ; unpaired Student's *t*-test; mean ± s.e.m.). Scale bars, 50 μm. **g**, **h**, PAS-stained ileal sections (**g**) with quantification of PAS<sup>+</sup> cells per total intestinal epithelial cell count along the crypt–villus axis shown in **h** ( $n = 5$ ; unpaired Student's *t*-test; mean ± s.e.m.). Scale bars, 50 μm.

**i**, Enteritis histology score of indicated genotypes ( $n = 3$ ; median shown; Kruskal–Wallis). **j**, Taqman qRT–PCR for MNV of faecal samples in three different animal facilities ( $n = 3$ ; mean ± s.e.m.). Note detectable MNV genome copy numbers in Boston (*Atg7/Xbp1<sup>ΔIEC</sup>*, Fig. 2c) and Innsbruck (*Atg16l1/Xbp1<sup>ΔIEC</sup>*, Extended Data Fig. 5g) colonies, and absence of MNV after re-derivation of colonies into the MNV-free Cambridge facility (*Atg16l1/Xbp1<sup>ΔIEC</sup>*, Fig. 2e), yet similar levels of enteritis in *Xbp1<sup>ΔIEC</sup>* and *Atg16l1/Xbp1<sup>ΔIEC</sup>* mice maintained in Innsbruck (MNV<sup>+</sup>) and Cambridge (MNV<sup>−</sup>) (Fig. 2b, e and Extended Data Fig. 5g) and relatively more severe disease in *Xbp1<sup>ΔIEC</sup>* mice from Boston (MNV<sup>+</sup>) (Fig. 2c). Sentinel mice from the Cambridge colony also tested negative for MNV by serological analysis (data not shown). These observations suggest that the inflammatory phenotypes observed are not necessarily dependent on MNV. \* $P < 0.05$ .





**Extended Data Figure 10 | Model of the interplay between UPR-induced autophagy, NF-κB signalling and inflammation.** **a**, Deletion of *Xbp1* in the intestinal epithelium, specifically in Paneth cells, leads to ER stress and activation of the PERK-eIF2α branch of the UPR. ATF4, a transcriptional mediator of this pathway, transactivates genes essential for autophagosome formation, such as *Map1lc3b* (*LC3b*) and *Atg7*, which catalyses the creation of the ATG12-ATG5 conjugate that stabilizes ATG16L1 through complex formation<sup>21</sup>. UPR-induced autophagy in the intestinal epithelium is essential for restoration of homeostasis and restraint of ER stress-induced intestinal inflammation due to XBP1 deficiency. Activation of the UPR in the setting of XBP1 deficiency results in activation of IRE1α, resulting in the recruitment of TRAF2 and activation of IKK2 leading to IκBα degradation<sup>4,27,45,46</sup>. As shown here, UPR-mediated autophagy serves an important role in restraining NF-κB activation, conceivably by removing hyperinflammatory ER membranes containing activated IRE1α. Pharmacological augmentation of this compensatory autophagy-dependent mechanism via inhibition of eIF2α dephosphorylation through salubrinal, or via the mTOR inhibitor rapamycin, results in amelioration of UPR-induced enteritis, which is driven by the commensal microbiota, NF-κB, and TNF-R1 signalling in intestinal epithelial cells and myeloid cells, whereby the ligand TNF may originate from XBP1-deficient intestinal epithelial cells<sup>4</sup>. **b**, ATG16L1 deficiency in intestinal epithelial cells leads to ER stress as revealed through upregulation of the chaperone GRP78 in intestinal epithelial cells, increased expression of GRP78

protein in Paneth cells, increased IRE1α expression and increased splicing of *Xbp1* mRNA in intestinal crypts as well as increased intestinal epithelial cell death. This leads to increased sensitivity of the epithelium to environmental triggers (for example, dextran sodium sulphate) that further challenge the UPR and its compensatory pathways. **c**, Deficiency of ATG16L1 or ATG7 in the intestinal epithelium results in abrogation of the compensatory autophagic mechanism that restrains IRE1α activity, conceivably via removal of hyperinflammatory ER membranes, and further fosters intestinal epithelial cell death in the context of ER stress due to *Xbp1* deficiency, resulting in spontaneous transmural small intestinal inflammation that is associated with further increases in NF-κB activation and cell death via the mechanisms described in **a**. The UPR allows for responses to a variety of signals that have an impact on protein folding, including genetic (for example, rare *XBP1* variants, *ORMDL3* as risk factor of inflammatory bowel disease<sup>4,47</sup>), environmental (for example, low O<sub>2</sub> tension in the intestinal tract) and microbial factors (for example, microbial toxins such as trierixin<sup>48</sup>), which determines the level of ER stress in the intestinal epithelium. UPR-induced autophagy function provides a buffer to cope with different levels of ER stress and vice versa. However, in the presence of genetic risk variants, such as *ATG16L1* (refs 1, 17, 49) or *IRGM* (ref. 50), which are relatively prevalent in the general population, this compensatory mechanism is impaired, resulting in development and/or exacerbation of intestinal inflammation in the setting of unabated ER stress.

# Antibody-mediated immunotherapy of macaques chronically infected with SHIV suppresses viraemia

Masashi Shingai<sup>1\*</sup>, Yoshiaki Nishimura<sup>1\*</sup>, Florian Klein<sup>2</sup>, Hugo Mouquet<sup>3</sup>, Olivia K. Donau<sup>1</sup>, Ronald Plishka<sup>1</sup>, Alicia Buckler-White<sup>1</sup>, Michael Seaman<sup>4</sup>, Michael Piatak Jr<sup>5</sup>, Jeffrey D. Lifson<sup>5</sup>, Dimitar S. Dimitrov<sup>6</sup>, Michel C. Nussenzweig<sup>2,6,7</sup> & Malcolm A. Martin<sup>1</sup>

Neutralizing antibodies can confer immunity to primate lentiviruses by blocking infection in macaque models of AIDS<sup>1–4</sup>. However, earlier studies of anti-human immunodeficiency virus type 1 (HIV-1) neutralizing antibodies administered to infected individuals or humanized mice reported poor control of virus replication and the rapid emergence of resistant variants<sup>5–7</sup>. A new generation of anti-HIV-1 monoclonal antibodies, possessing extraordinary potency and breadth of neutralizing activity, has recently been isolated from infected individuals<sup>8</sup>. These neutralizing antibodies target different regions of the HIV-1 envelope glycoprotein including the CD4-binding site, glycans located in the V1/V2, V3 and V4 regions, and the membrane proximal external region of gp41 (refs 9–14). Here we have examined two of the new antibodies, directed to the CD4-binding site and the V3 region (3BNC117 and 10-1074, respectively), for their ability to block infection and suppress viraemia in macaques infected with the R5 tropic simian–human immunodeficiency virus (SHIV)-AD8, which emulates many of the pathogenic and immunogenic properties of HIV-1 during infections of rhesus macaques<sup>15,16</sup>. Either antibody alone can potently block virus acquisition. When administered individually to recently infected macaques, the 10-1074 antibody caused a rapid decline in virus load to undetectable levels for 4–7 days, followed by virus rebound during which neutralization-resistant variants became detectable. When administered together, a single treatment rapidly suppressed plasma viraemia for 3–5 weeks in some long-term chronically SHIV-infected animals with low CD4<sup>+</sup> T-cell levels. A second cycle of anti-HIV-1 monoclonal antibody therapy, administered to two previously treated animals, successfully controlled virus rebound. These results indicate that immunotherapy or a combination of immunotherapy plus conventional antiretroviral drugs might be useful as a treatment for chronically HIV-1-infected individuals experiencing immune dysfunction.

SHIV-AD8 was selected as challenge virus for this study because several clinical features observed during infections of macaques were similar to those reported in HIV-1-infected individuals. SHIV-AD8 consistently establishes sustained set-point viraemia in macaques inoculated by the intravenous or intrarectal routes and causes unrelenting depletion of CD4<sup>+</sup> T lymphocytes<sup>15,16</sup>. During the acute infection, SHIV-AD8 targets memory CD4<sup>+</sup> T cells in blood and at effector sites in tissues. Their gradual depletion is subsequently followed by the loss of the naive CD4<sup>+</sup> T lymphocyte subset. The latter heralds the onset of symptomatic immunodeficiency in macaques characterized by the development of opportunistic infections (*Mycobacterium*, *Pneumocystis*, *Cryptosporidium* species), lymphomas, marked weight loss, and death within 2–4 years of virus inoculation. In addition, SHIV-AD8-infected macaques generate cross-reactive antibodies, capable of neutralizing

tier 1 and tier 2 HIV-1 isolates, including one ‘elite neutralizer’ macaque producing potent cross-clade neutralizing activity<sup>8,17,18</sup>. As is frequently the case for HIV-1 elite neutralizers, resistant variants emerged in this SHIV-AD8-infected animal, which succumbed to AIDS at week 117 after infection<sup>19</sup>.

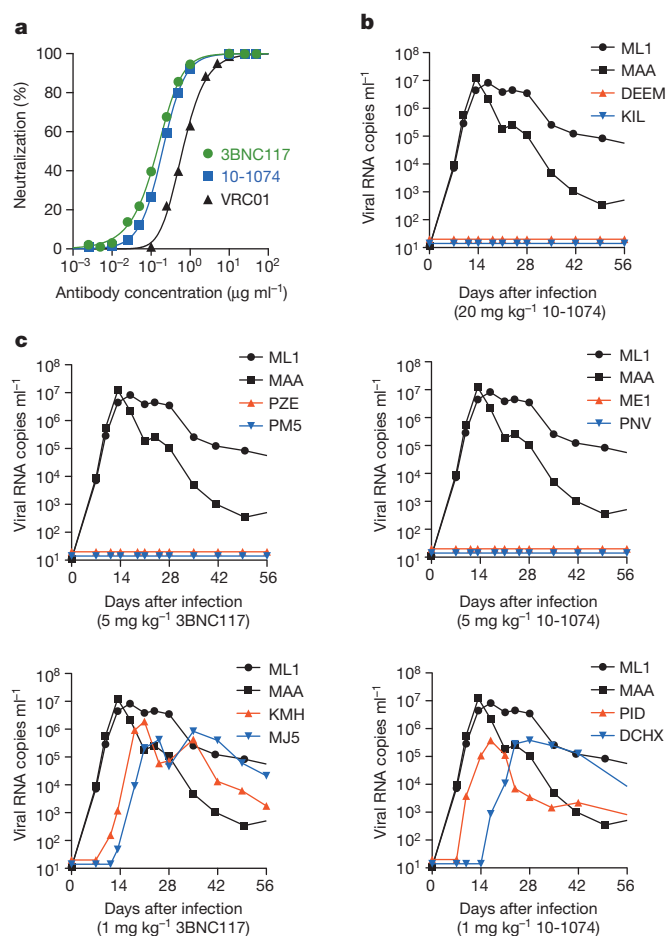
The neutralization sensitivities of SHIV-AD8EO (ref. 17)—a molecularly cloned derivative of SHIV-AD8 (ref. 16)—to the 10-1074 (ref. 20), 3BNC117 (ref. 21) and VRC01 (ref. 22) monoclonal antibodies were measured in the TZM-bl cell assay (Fig. 1a). The half-maximum inhibitory concentration (IC<sub>50</sub>) values determined for 10-1074, 3BNC117 and VRC01 against SHIV-AD8EO were 0.20 µg ml<sup>−1</sup>, 0.14 µg ml<sup>−1</sup> and 0.63 µg ml<sup>−1</sup>, respectively, indicating that the 10-1074 and 3BNC117 monoclonal antibodies had similar activities and either one was more potent than VRC01 against SHIV-AD8EO *in vitro*.

A more critical test of neutralization efficacy is the prevention of virus acquisition *in vivo* after passive transfer of neutralizing antibodies and subsequent virus challenge. This was examined by administering the 10-1074 or 3BNC117 monoclonal antibodies 24 h before an intrarectal challenge of Indian origin rhesus macaques with 1,000 tissue infectious dose<sub>50</sub> (TCID<sub>50</sub>) (approximately 3 animal infectious doses<sub>50</sub> (AID<sub>50</sub>)<sup>15</sup>), an inoculum size we have previously determined is sufficient to establish SHIV-AD8EO infections *in vivo* after a single inoculation by this route in 10 of 10 macaques. As shown in Fig. 1b, the transfer of the 10-1074 monoclonal antibody at a dose of 20 mg kg<sup>−1</sup> or 5 mg kg<sup>−1</sup> to macaques prevented virus acquisition in 2 of 2 and 2 of 2 macaques, respectively. The administration of 1 mg kg<sup>−1</sup> of 10-1074, however, failed to protect either of two animals. For the 3BNC117 passive transfer, the monoclonal antibody *in vivo* titration was initiated at a dose of 5 mg kg<sup>−1</sup>, which blocked virus acquisition in 2 of 2 macaques (Fig. 1c). In contrast, 2 of 2 macaques became infected when the dose of 3BNC117 was reduced to 1 mg kg<sup>−1</sup>. The plasma concentrations at the time of challenge for both monoclonal antibodies were comparable (approximately 100 µg ml<sup>−1</sup>) in the four macaques treated with a dose of 5 mg kg<sup>−1</sup>.

Recent studies have reported that combinations of three or more of the newly cloned broad and potent antibodies effectively suppressed HIV-1 viraemia in humanized mice<sup>10,23,24</sup>. However, humanized mice carry a lower viral load, and do not have an intact adaptive or innate immune system. We have also previously reported that in the presence of neutralizing antibody, the clearance of circulating HIV-1 *in vivo* is accelerated from approximately 20 min to 3 or 4 min<sup>25</sup>. The potential therapeutic benefit of anti-HIV-1 neutralizing antibodies controlling virus replication in infected macaques was evaluated by administering either 10-1074 or 3BNC117 alone (monotherapy) to two animals, 12 weeks after virus inoculation, when post-peak set-point viraemia levels had been established. On the basis of the pre-exposure prevention experiments

<sup>1</sup>Laboratory of Molecular Microbiology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>2</sup>Laboratory of Molecular Immunology, The Rockefeller University, New York, New York 10065, USA. <sup>3</sup>Laboratory of Humoral Response to Pathogens, Department of Immunology, Institut Pasteur, 75015 Paris, France. <sup>4</sup>Center for Virology and Vaccine Research, Beth Israel Deaconess Medical Center, Harvard Medical School, 3 Blackfan Circle, E/CLS-1001 Boston, Massachusetts 02115, USA. <sup>5</sup>AIDS and Cancer Virus Program, SAIC-Frederick, Inc., Frederick National Laboratory for Cancer Research, Frederick, Maryland 21702, USA. <sup>6</sup>Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>7</sup>Howard Hughes Medical Institute, The Rockefeller University, New York, New York 10065, USA.

\*These authors contributed equally to this work.



**Figure 1 | HIV monoclonal antibodies block SHIV acquisition.**

**a**, Neutralization of SHIV-AD8EO by three anti-HIV-1 monoclonal antibodies assayed in TZM-bl cells. **b**, Pre-exposure passive transfer of 10-1074 monoclonal antibody to macaques followed by SHIV-AD8EO intrarectal challenge ( $20 \text{ mg kg}^{-1}$  (top),  $5 \text{ mg kg}^{-1}$  (middle) and  $1 \text{ mg kg}^{-1}$  (bottom)). **c**, Pre-exposure passive transfer of 3BNC117 monoclonal antibody to macaques followed by SHIV-AD8EO intrarectal challenge. Macaques ML1 and MAA received  $20 \text{ mg kg}^{-1}$  of control anti-dengue virus NS1 IgG1 monoclonal antibody.

described above, infected macaques were treated with either antibody at a dose of  $10 \text{ mg kg}^{-1}$  (Table 1). Both recipients of the 3BNC117 monoclonal antibody experienced rapid declines of plasma viraemia to background levels at day 10 after the start of treatment (Fig. 2a). Virus rebound became detectable on day 20 in these two macaques. The 10-1074-treated animal MCN sustained an unexpectedly rapid and a greater than  $10^3$  reduction of plasma viral loads to undetectable levels by day 6 of treatment initiation. However, the effects of 10-1074 monoclonal antibody administration were more variable and of short duration

(4 days) (Fig. 2a and Extended Data Fig. 1). Furthermore, single genome amplification (SGA) analysis of virus in the four monotherapy recipients revealed that the rebound virus present in both of the 10-1074-treated macaques had sustained changes that eliminated the gp120 Asn 332 glycan, rendering the virus resistant to this antibody (Extended Data Fig. 2)<sup>10</sup>.

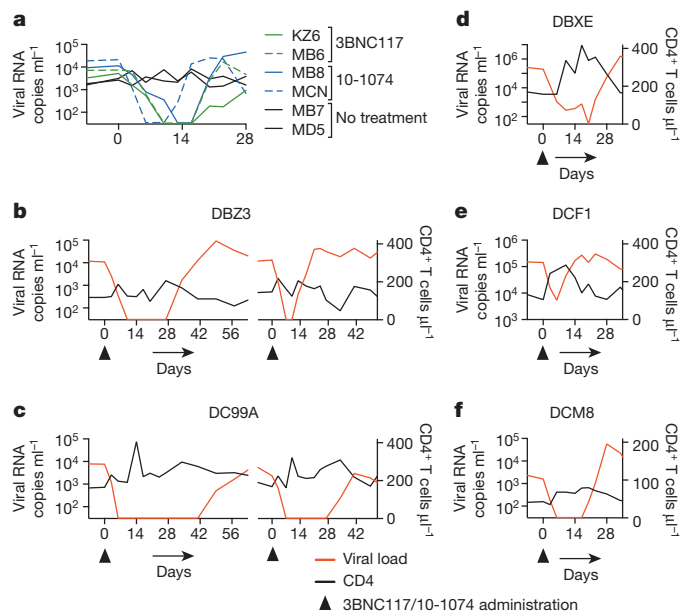
The known failure of antiretroviral drug monotherapy to durably control HIV-1 and SIV replication *in vivo* prompted us to co-administer 3BNC117 and 10-1074 monoclonal antibodies, each at a dose of  $10 \text{ mg kg}^{-1}$ , to chronically SHIV-AD8EO-infected macaques. In addition, the capacity of the molecularly cloned SHIV-AD8EO to cause unrelenting  $\text{CD4}^+$  T-cell depletion and symptomatic immunodeficiency provided the opportunity to assess the potential therapeutic effects of combination anti-HIV-1 neutralizing monoclonal antibodies in two groups of infected macaques experiencing the pathogenic effects of SHIV-AD8EO infection. The first group consisted of two clinically asymptomatic animals (DBZ3 and DC99A) that had been infected for 159 weeks and had sustained similar and significant declines of circulating and bronchoalveolar lavage (BAL)  $\text{CD4}^+$  T cells (Table 1 and Extended Data Figs 3 and 4). At the time of monoclonal antibody administration, the plasma viral loads in macaques DBZ3 and DC99A were  $1.08 \times 10^4$  and  $7.6 \times 10^3$  RNA copies  $\text{ml}^{-1}$ , respectively. Both macaques responded to combination anti-HIV-1 monoclonal antibody treatment with immediate and rapid reductions of plasma viraemia to undetectable levels within 7–10 days (Fig. 2b, c). Suppression of measurable SHIV-AD8EO in the plasma of macaques DBZ3 and DC99A, after a single administration of the two monoclonal antibodies, lasted 18 and 36 days, respectively, exceeding the 4–7 day window of virus suppression observed during monotherapy. In each case, plasma viraemia rebounded to pre-treatment levels. A second cycle of combination immunotherapy was administered to macaques DBZ3 and DC99A to determine whether the re-emergence of detectable plasma viraemia simply reflected insufficient levels of circulating antibodies in these animals. As shown in the right panels of Figs 2b, c, the viral loads in each animal became undetectable by day 7 of the second treatment cycle. Viraemia was suppressed for 4 days in macaque DBZ3 and 28 days in macaque DC99A. SGA analysis demonstrated that the rebound virus that emerged in macaque DC99A, after the second round of immunotherapy, carried a gp120 Asn332Asp mutation, indicating that resistance to the 10-1074 component of the antibody combination had occurred (Extended Data Fig. 5b).

Combination monoclonal antibody therapy was also evaluated in a second group of three chronically infected animals (DBX3, DCF1 and DCM8), which had also been infected for more than 3 years but were clinically symptomatic, experiencing intermittent diarrhoea and or anorexia (Table 1). At the time of monoclonal antibody administration, the level of circulating  $\text{CD4}^+$  T cells in one of these macaques (DCM8) was only  $43 \text{ cells } \mu\text{l}^{-1}$  and was higher in animals DCF1 ( $105 \text{ cells } \mu\text{l}^{-1}$ ) and DBX3 ( $158 \text{ cells } \mu\text{l}^{-1}$ ). Plasma viral loads exceeded  $10^5$  RNA copies  $\text{ml}^{-1}$  in animals DBX3 and DCF1 and were significantly lower ( $1.59 \times 10^3$  RNA copies  $\text{ml}^{-1}$ ) in macaque DCM8. As shown in Fig. 2d–f and Extended Data Fig. 3, the administration of

**Table 1 | SHIV-AD8-infected macaques treated with anti-HIV monoclonal antibodies**

Animal	Pre-infection	Treatment initiation (weeks after infection)	Pre-monoclonal antibody treatment		Clinical status	Monoclonal antibody administered	
	( $\text{CD4}^+$ T cells ( $\text{cells } \mu\text{l}^{-1}$ ))		$\text{CD4}^+$ T cells ( $\text{cells } \mu\text{l}^{-1}$ )	Viral load (RNA copies $\text{ml}^{-1}$ )		3BNC117 ( $\text{mg kg}^{-1}$ )	10-1074 ( $\text{mg kg}^{-1}$ )
KZ6	881	12	1,447	$5.22 \times 10^3$	Post-acute set point	10	0
MB6	545	12	1,598	$7.43 \times 10^3$	Post-acute set point	10	0
MB8	1,512	12	2,625	$1.14 \times 10^4$	Post-acute set point	0	10
MCN	210	12	599	$2.14 \times 10^4$	Post-acute set point	0	10
DBZ3	650	159	118	$1.08 \times 10^4$	Asymptomatic	10	10
DC99A	623	159	165	$7.60 \times 10^3$	Asymptomatic	10	10
DBX3	1,585	163	158	$1.96 \times 10^5$	Intermittent diarrhoea	10	10
DCF1	1,203	157	105	$1.44 \times 10^5$	Intermittent diarrhoea	10	10
DCM8	608	163	43	$1.59 \times 10^3$	Intermittent diarrhoea	10	10





**Figure 2 | Suppression of plasma viraemia after monotherapy or combination anti-HIV-1 neutralizing antibody treatment.** **a**, Plasma viral loads in post-acute set-point SHIV-infected rhesus macaques with/without single monoclonal antibody treatment. **b–f**, Plasma viral loads and total CD4<sup>+</sup> T-cell numbers in chronically SHIV-infected rhesus macaques (DBZ3 (**b**), DC99A (**c**), DBXE (**d**), DCF1 (**e**) and DCM8 (**f**)) after combination monoclonal antibody treatment.

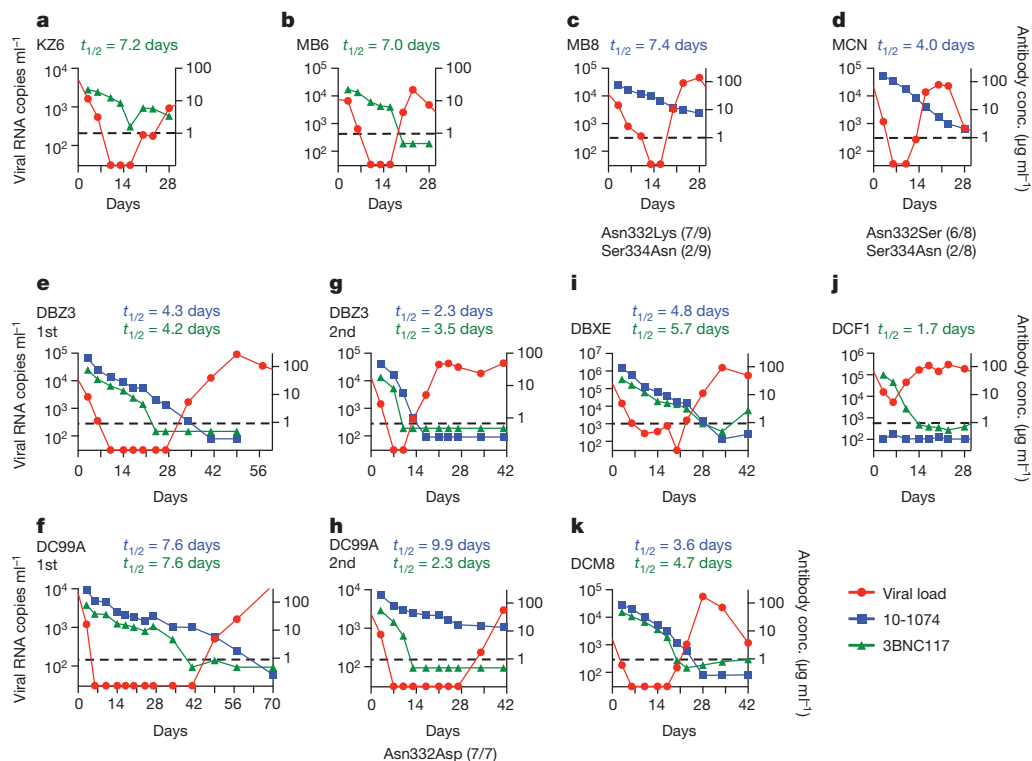
the two monoclonal antibodies to macaque DBXE resulted in reduction of viraemia from  $2.0 \times 10^5$  RNA copies at day 0 to undetectable levels in plasma at day 20. This was followed, within a few days, by a resurgence of high levels of circulating virus in DBXE. Macaque DCM8, with more modest plasma virus loads and very low numbers of circulating CD4<sup>+</sup> T cells, experienced a rapid decline of viraemia

to undetectable levels between days 6 and 17 after the initiation of monoclonal antibody treatment. Finally, animal DCF1, previously reported to have generated broadly reacting anti-HIV-1 neutralizing antibodies<sup>17</sup>, exhibited a transient and a comparatively modest 27-fold reduction of plasma viraemia by day 6 in response to combination monoclonal antibody therapy, before the viral loads returned to high pre-treatment levels.

SGA analysis of emerging virus populations in the five long-term chronically infected macaques revealed no changes affecting sensitivity to either 10-1074 or 3BNC117 except for the rebounding virus after the second treatment cycle of macaque DC99A noted above (Extended Data Fig. 5).

Because some of the recipients of combination antibody treatment had prolonged control of detectable plasma viraemia, peripheral blood mononuclear cell (PBMC)-associated viral RNA and DNA levels were evaluated before and after monoclonal antibody administration. As shown in Extended Data Table 1, monoclonal antibody treatment resulted in reduced levels of cell-associated viral RNA in each animal contemporaneous with the suppression of its plasma viraemia. No consistent change in cell-associated viral DNA levels occurred as a result of antibody treatment.

Co-administration of neutralizing monoclonal antibodies to chronically SHIV-AD8EO-infected macaques resulted in a transient elevation of circulating CD4<sup>+</sup> T-cell levels, particularly in late-stage animals with very high viral loads. The CD4<sup>+</sup> T-cell numbers in macaques DBXE and DCF1 increased 2–3-fold during the period of monoclonal-antibody-mediated virus suppression, but gradually declined to pre-treatment levels as viraemia again became detectable (Extended Data Fig. 6). The corresponding elevations and declines of both the naive and memory CD4<sup>+</sup> T-cell subsets in animals DBXE and DCF1 would be consistent with an antibody-mediated tissue redistribution mechanism, given that the R5 tropic SHIV-AD8EO exclusively targets memory, not naive, CD4<sup>+</sup> T lymphocytes. The levels of memory CD4<sup>+</sup> T cells at an effector site (from BAL) did not change appreciably during immunotherapy (Extended Data Fig. 4).



**Figure 3 | Plasma viraemia rebounds in SHIV-infected macaques when neutralizing antibody levels decline.** **a–k**, Viral RNA levels and concentrations of 10-1074 or 3BNC117 monoclonal antibodies in plasma at

various times after initiation of single or combination antibody treatment. In panels **c**, **d**, **h**, the gp120 changes present in the rebound virus populations and their frequencies are indicated at the bottom.

Plasma concentrations of each monoclonal antibody were determined by measuring the neutralizing activity in plasma against selected HIV-1 pseudovirus strains sensitive to one or the other, but not to both, antibodies (Extended Data Fig. 7). In treated animals, suppression of SHIV-AD8EO viraemia was maintained until a threshold plasma monoclonal antibody concentration of approximately  $5 \mu\text{g ml}^{-1}$  was reached, except when escape variants emerged in animals MB8, MCN and the second cycle of treatment in macaque DC99A (Fig. 3 and Extended Data Fig. 8). This was even the case for macaque DCF1, for which only a modest and transient reduction of plasma viral RNA levels was observed after combination therapy. Notably, the monoclonal antibodies administered to this clinically symptomatic macaque had a shortened half-life or were undetectable (Fig. 3j). In this regard, macaque DCF1 had to be euthanized on day 56 after treatment initiation and a necropsy revealed severe enteropathy, characterized by disseminated gastrointestinal cryptosporidiosis, pancreatitis and cholangitis, perhaps explaining the rapid loss of neutralizing antibodies in this animal.

The decay rate constants and the corresponding half-lives of plasma viral RNA after monotherapy or combination neutralizing antibody treatments were calculated from the initial slopes of the HIV-1 plasma RNA decline (Extended Data Table 2). Furthermore, unlike pharmacological antiretroviral therapy (ART) in which two phases of viral decay lasting several weeks is commonly seen, treatment with neutralizing antibodies induced a single phase of exponential decline ( $t_{1/2} = 1.13$  days). For the macaques (DC99A and DBZ3) that received two cycles of treatment, the decay rate constant for the second treatment was higher than that for the first treatment for one of the animals and lower for the other (Extended Data Table 2).

Our findings demonstrate that combination anti-HIV neutralizing monoclonal antibody treatment is superior to monotherapy and can rapidly and potently suppress plasma viraemia in chronically SHIV-infected macaques with low  $\text{CD4}^+$  T-cell levels and symptomatic disease. Animals administered a single neutralizing monoclonal antibody experienced a relatively short period of plasma viral RNA suppression, and neutralization-resistant progeny virus emerged in some macaques, attesting to the robustness of SHIV-AD8EO *in vivo*. Our results further indicate that administration of monoclonal antibodies, alone or in combination with currently available antiretroviral drug regimens<sup>24</sup>, could be useful therapeutically, particularly for chronically infected individuals with compromised immune systems.

## METHODS SUMMARY

The neutralization activities of the broadly acting 3BNC117 (ref. 21)  $\text{CD4}$ -binding site and 10-1074 (ref. 20) gp120 Asn-332-dependent neutralizing monoclonal antibodies against SHIV-AD8EO were initially determined in the TZM-bl cell system against SHIV-AD8EO as previously described<sup>15,17</sup>. Their capacities to block virus acquisition or to control plasma viraemia in chronically infected animals challenged with the R5-tropic SHIV-AD8EO were assessed by monitoring plasma viral loads and cell-associated viral nucleic acids; levels of  $\text{CD4}^+$  T-cell subsets were measured by flow cytometry as previously reported<sup>16</sup>. SGA analyses of circulating viral variants and the determination of antibody levels in plasma have been previously described<sup>26</sup>. Plasma concentration of neutralizing antibodies was determined by measuring neutralizing activity against HIV-1 pseudovirus preparations only susceptible to either 10-1074 or 3BNC117.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 5 August; accepted 11 October 2013.**

**Published online 30 October; corrected online 13 November 2013 (see full-text HTML version for details).**

1. Mascola, J. R. *et al.* Protection of macaques against pathogenic simian/human immunodeficiency virus 89.6PD by passive transfer of neutralizing antibodies. *J. Virol.* **73**, 4009–4018 (1999).
2. Moldt, B. *et al.* Highly potent HIV-specific antibody neutralization *in vitro* translates into effective protection against mucosal SHIV challenge *in vivo*. *Proc. Natl Acad. Sci. USA* **109**, 18921–18925 (2012).

3. Nishimura, Y. *et al.* Determination of a statistically valid neutralization titer in plasma that confers protection against simian-human immunodeficiency virus challenge following passive transfer of high-titered neutralizing antibodies. *J. Virol.* **76**, 2123–2130 (2002).
4. Parren, P. W. *et al.* Antibody protects macaques against vaginal challenge with a pathogenic R5 simian/human immunodeficiency virus at serum levels giving complete neutralization *in vitro*. *J. Virol.* **75**, 8340–8347 (2001).
5. Mehndru, S. *et al.* Adjunctive passive immunotherapy in human immunodeficiency virus type 1-infected individuals treated with antiviral therapy during acute and early infection. *J. Virol.* **81**, 11016–11031 (2007).
6. Pognard, P. *et al.* Neutralizing antibodies have limited effects on the control of established HIV-1 infection *in vivo*. *Immunity* **10**, 431–438 (1999).
7. Trkola, A. *et al.* Delay of HIV-1 rebound after cessation of antiretroviral therapy through passive transfer of human neutralizing antibodies. *Nature Med.* **11**, 615–622 (2005).
8. Burton, D. R. *et al.* A blueprint for HIV vaccine discovery. *Cell Host Microbe* **12**, 396–407 (2012).
9. Huang, J. *et al.* Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature* **491**, 406–412 (2012).
10. Klein, F. *et al.* HIV therapy by a combination of broadly neutralizing antibodies in humanized mice. *Nature* **492**, 118–122 (2012).
11. Kong, L. *et al.* Supersite of immune vulnerability on the glycosylated face of HIV-1 envelope glycoprotein gp120. *Nature Struct. Mol. Biol.* **20**, 796–803 (2013).
12. Walker, L. M. *et al.* Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* **477**, 466–470 (2011).
13. Walker, L. M. *et al.* Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science* **326**, 285–289 (2009).
14. Wu, X. *et al.* Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* **329**, 856–861 (2010).
15. Gautam, R. *et al.* Pathogenicity and mucosal transmissibility of the R5-tropic simian/human immunodeficiency virus SHIV(AD8) in rhesus macaques: implications for use in vaccine studies. *J. Virol.* **86**, 8516–8526 (2012).
16. Nishimura, Y. *et al.* Generation of the pathogenic R5-tropic simian/human immunodeficiency virus SHIVAD8 by serial passaging in rhesus macaques. *J. Virol.* **84**, 4769–4781 (2010).
17. Shingai, M. *et al.* Most rhesus macaques infected with the CCR5-tropic SHIV(AD8) generate cross-reactive antibodies that neutralize multiple HIV-1 strains. *Proc. Natl Acad. Sci. USA* **109**, 19769–19774 (2012).
18. Walker, L. M. *et al.* Rapid development of glycan-specific, broad, and potent anti-HIV-1 gp120 neutralizing antibodies in an R5 SIV/HIV chimeric virus infected macaque. *Proc. Natl Acad. Sci. USA* **108**, 20125–20129 (2011).
19. Sadjadpour, R. *et al.* Emergence of gp120 V3 variants confers neutralization resistance in an R5 simian-human immunodeficiency virus-infected macaque elite neutralizer that targets the N332 glycan of the human immunodeficiency virus type 1 envelope glycoprotein. *J. Virol.* **87**, 8798–8804 (2013).
20. Mouquet, H. *et al.* Complex-type N-glycan recognition by potent broadly neutralizing HIV antibodies. *Proc. Natl Acad. Sci. USA* **109**, E3268–E3277 (2012).
21. Scheid, J. F. *et al.* Sequence and structural convergence of broad and potent HIV antibodies that mimic  $\text{CD4}$  binding. *Science* **333**, 1633–1637 (2011).
22. Zhou, T. *et al.* Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science* **329**, 811–817 (2010).
23. Diskin, R. *et al.* Restricting HIV-1 pathways for escape using rationally designed anti-HIV-1 antibodies. *J. Exp. Med.* **210**, 1235–1249 (2013).
24. Horwitz, J. A. *et al.* HIV-1 suppression and durable control by combining single broadly neutralizing antibodies and antiretroviral drugs in humanized mice. *Proc. Natl Acad. Sci. USA*. <http://dx.doi.org/10.1073/pnas.1315295110> (2013).
25. Igarashi, T. *et al.* Human immunodeficiency virus type 1 neutralizing antibodies accelerate clearance of cell-free virions from blood plasma. *Nature Med.* **5**, 211–216 (1999).
26. Keele, B. F. *et al.* Low-dose rectal inoculation of rhesus macaques by SIVsmE660 or SIVmac251 recapitulates human mucosal infection by HIV-1. *J. Exp. Med.* **206**, 1117–1134 (2009).

**Acknowledgements** We thank K. Tomioka and R. Kruthers for determining plasma viral RNA loads and B. Skopets, W. Magnanelli and R. Petros for diligently assisting in the maintenance of animals and assisting with procedures. We also thank D. R. Burton, The Scripps Institute, for providing anti-dengue virus neutralizing monoclonal antibody (DEN-3). This work was supported by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases, National Institutes of Health (NIH) and, in part, with federal funds from the National Cancer Institute, NIH, under contract HHSN261200800001E.

**Author Contributions** M.S., Y.N., M.C.N. and M.A.M. designed the experiments; M.S., Y.N., F.K., H.M., O.K.D., R.P., A.B.-W. and M.P. performed the experiments; M.S., Y.N., F.K., M.P., J.D.L., D.D., M.C.N. and M.A.M. analysed the data; and M.S., Y.N., M.C.N. and M.A.M. wrote the manuscript. M.S. and Y.N. contributed equally to the work.

**Author Information** The SHIV-AD8 gp120 sequences known to confer resistance to the 10-1074 or 3BNC117 monoclonal antibodies have been deposited in GenBank/EMBL/DBJ under accession numbers KF738375 to KF738446. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.C.N. ([nussen@mail.rockefeller.edu](mailto:nussen@mail.rockefeller.edu)) or M.A.M. ([malm@nih.gov](mailto:malm@nih.gov)).

## METHODS

**Virus.** The origin and preparation of the tissue-culture-derived SHIV-AD8EO stock have been previously described<sup>17</sup>.

**Animal experiments.** Twenty-three male and female rhesus macaques (*Macaca mulatta*) of Indian genetic origin ranging from 2 to 10 years of age were maintained in accordance with Guide for the Care and Use of Laboratory Animals Report no. NIH 85-23 (Department of Health and Human Services, Bethesda, Maryland, 1985) and were housed in a biosafety level 2 NIAID facility. Phlebotomies, euthanasia and sample collection were performed as previously described<sup>27</sup>. All animals were negative for the MHC class I *Mamu-A\*01*, *Mamu-B\*08* and *Mamu-B\*17* allele.

**Antibodies.** Two monoclonal antibodies 3BNC117 (ref. 21) and 10-1074 (ref. 20) were isolated and produced as described elsewhere. DEN3, a dengue virus NS1-specific human IgG1 antibody<sup>2</sup>, was used as the negative control antibodies in this study. All antibodies were administered intravenously.

**Quantification of viral nucleic acids.** Viral RNA levels in plasma were determined by real-time reverse transcription-PCR (ABI Prism 7900HT sequence detection system; Applied Biosystems) as previously reported<sup>27</sup>. Ultrasensitive measurement of cell-associated SIV RNA and DNA were determined by a nested, hybrid real-time/digital PCR assay<sup>28</sup>.

**Neutralization assays.** The *in vitro* potency of each monoclonal antibody was assessed by TZM-bl entry assay with pseudotyped SHIV-AD8EO as previously reported<sup>15,17</sup>.

**Lymphocyte immunophenotyping.** EDTA-treated blood samples were stained for flow cytometric analysis as previously described<sup>16</sup>.

**Antibody concentrations and  $t_{1/2}$  estimations in plasma.** Plasma concentrations of 10-1074 and 3BNC117 neutralization activity were separately determined against HIV-1 virus strains that are sensitive to one but not the other monoclonal

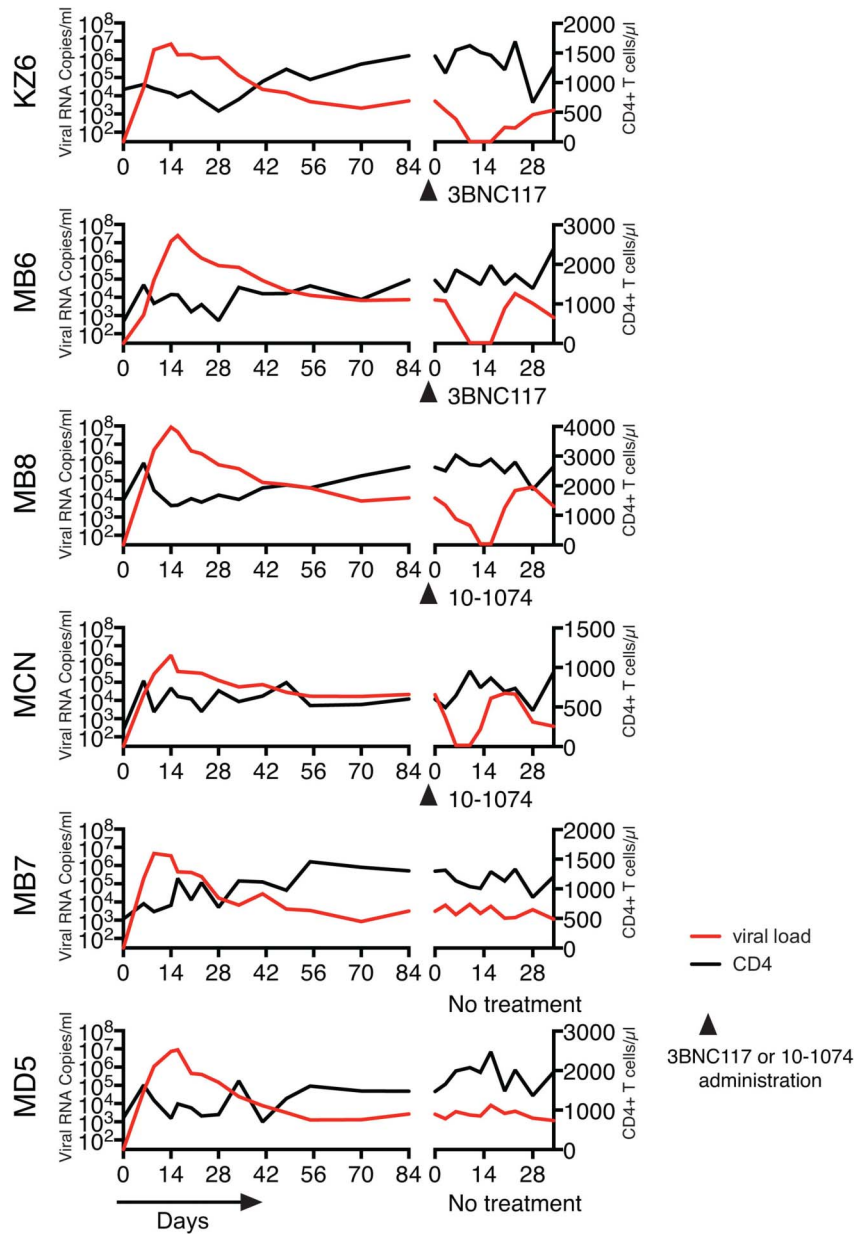
antibody as well as not, or very weakly, sensitive to autologous antibodies in macaque plasma. All samples were heat-inactivated for 1 h at 56 °C and neutralizing activity was measured by using a TZM-bl assay as previously described<sup>29</sup>. The half-maximum inhibitory dose (ID<sub>50</sub>) titres were determined by serial plasma dilutions starting at 1:20; antibody concentrations were calculated by multiplying each ID<sub>50</sub> value.  $t_{1/2}$  of plasma monoclonal antibodies was estimated by  $\ln(2)/k$ .

**Single genomic analysis.** Single-genome amplification of full-length HIV-1 *env* genes was performed as reported<sup>26</sup>.

**Determination of plasma viral RNA decay.** The kinetics of plasma HIV-1 RNA after the start of therapy was analysed as previously described<sup>30</sup> by a simple one-exponential model:  $\ln(V) = c - kt$ , where  $V$  is the HIV-1 RNA concentration, and  $k$  is virus decay rate constant (note that here  $k$  corresponds to  $k_1$  (ref. 30) and  $c$  = a constant close to the  $\ln$  of the baseline virus concentration). The half-life  $t_{1/2}$  was calculated as  $0.693/k$ . The data were fitted using the program Excel and the goodness of fit was estimated by the correlation coefficients, which were in the range from 0.91 to 1.0. All time points in the exponential phase of decline were used for calculation.

27. Endo, Y. *et al.* Short- and long-term clinical outcomes in rhesus monkeys inoculated with a highly pathogenic chimeric simian/human immunodeficiency virus. *J. Virol.* **74**, 6935–6945 (2000).
28. Hansen, S. G. *et al.* Profound early control of highly pathogenic SIV by an effector memory T-cell vaccine. *Nature* **473**, 523–527 (2011).
29. Montefiori, D. C. Evaluating neutralizing antibodies against HIV, SIV, and SHIV in luciferase reporter gene assays. *Curr. Protocols Immunol.* Ch. 12, Unit 12.11 (2005).
30. Polis, M. A. *et al.* Correlation between reduction in plasma HIV-1 RNA concentration 1 week after start of antiretroviral treatment and longer-term efficacy. *Lancet* **358**, 1760–1765 (2001).





**Extended Data Figure 1 | Treatment of SHIV-infected macaques with single anti-HIV-1 neutralizing monoclonal antibodies.** Plasma viral loads and total CD4<sup>+</sup> T cell numbers before (the initial 84 days of the SHIV-AD8EO infection) and during single monoclonal antibody treatment are shown. KZ6 and MB6

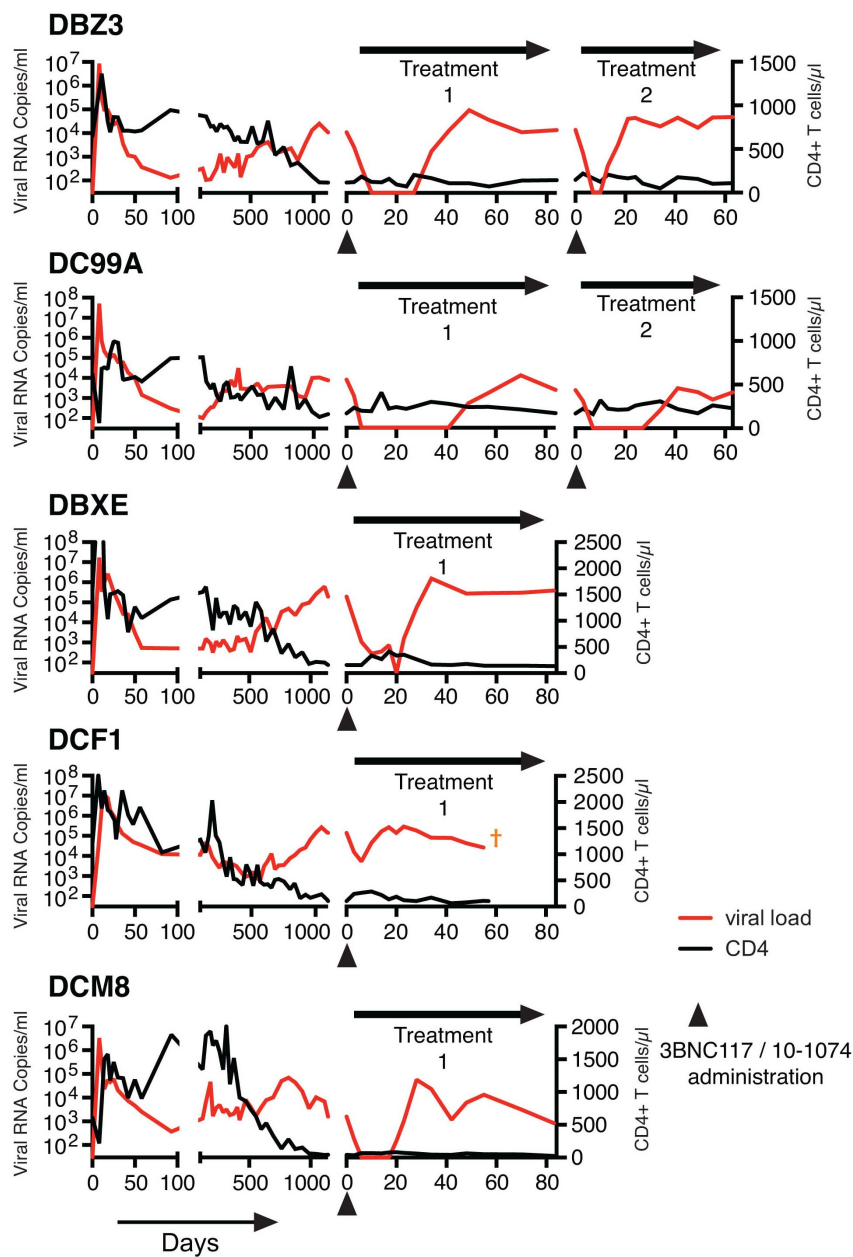
received the 3BNC117 monoclonal antibody and MB8 and MCN were administered the 10-1074 monoclonal antibody. Macaques MB7 and MD5 were not treated.

## Monotherapy

		320	332	335	276	284	364	372	424	434	454	460	468	478
		GDIIGDIRQAHNCNISR			SNF <sup>332</sup> TDNTKNI		SSGGDPEIV		INMQEVGKAM		LTRDGGN		FRPGGGDMRDN	
3BNC117	KZ6	*****			*****		*****		*****		*****		*****	
		*****			*****		*****		*****		*****		*****	
		*****			*****		*****		*****		*****		*****	
		*****			*****		*****		*****		*****		*****	
		*****			*****		*****		*****		*****		*****	
	MB6	*****			*****		*****		*****		*****		*****	
		*****			*****		*****		*****		*****		*****	
		*****			*****		*****		*****		*****		*****	
		*****			*****		*****		*****		*****		*****	
		*****			*****		*****		*****		*****		*****	
10-1074	MB8	*****	332		*****		*****		*****		*****		*****	
		*****	K		*****	M	*****		*****		*****		*****	
		*****	K		*****	M	*****		*****		*****		*****	
		*****	K		*****	M	*****		*****		*****		*****	
		*****	K		*****	M	*****		*****		*****		*****	
	MCN	*****	K		*****		*****		*****		*****		*****	
		*****	N		*****		*****		*****		*****		*****	
		*****	N		*****	R	*****		*****	I	*****		*****	
		*****	S		*****		*****		*****		*****		*****	
		*****	S		*****		*****		*****		*****		*****	
		*****	S		*****		*****		*****		*****		*****	
		*****	S		*****		*****		*****		*****		*****	
		*****	S		*****		*****		*****		*****		*****	
		*****	S		*****		*****		*****		*****		*****	
		*****	N		*****		*****		*****		*****		*****	
		*****	N		*****		*****		*****		*****		*****	D

**Extended Data Figure 2 | SGA analysis of selected SHIV-AD8EO gp120 sequences, present in rebound virus after single monoclonal antibody immunotherapy, and known to confer resistance to 10-1074 or 3BNC117 monoclonal antibody.** SGA was used to amplify plasma viral RNA after monoclonal antibody treatment from the plasma of animals KZ6 (day 28

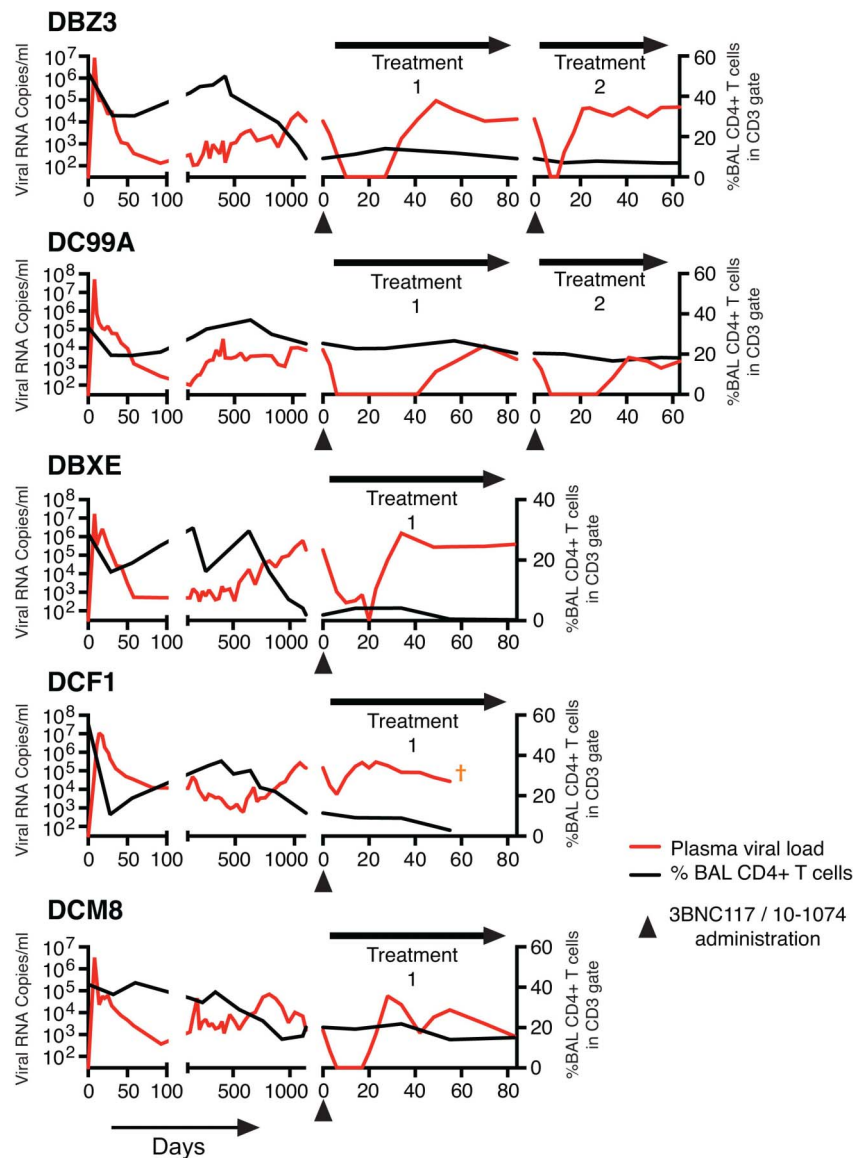
( $n = 8$ )), MB6 (day 23 ( $n = 8$ )), MB8 (day 23 ( $n = 9$ )) and MCN (day 23 ( $n = 8$ )). The gp120 sequences at the top are present in the SHIV-AD8EO molecular clone inoculum. Mutations conferring resistance are highlighted in red.



**Extended Data Figure 3 | Circulating CD4<sup>+</sup> T cells in five chronically SHIV-infected macaques treated with two anti-HIV-1 neutralizing monoclonal antibodies.** Plasma viral loads and total CD4<sup>+</sup> T-cell numbers

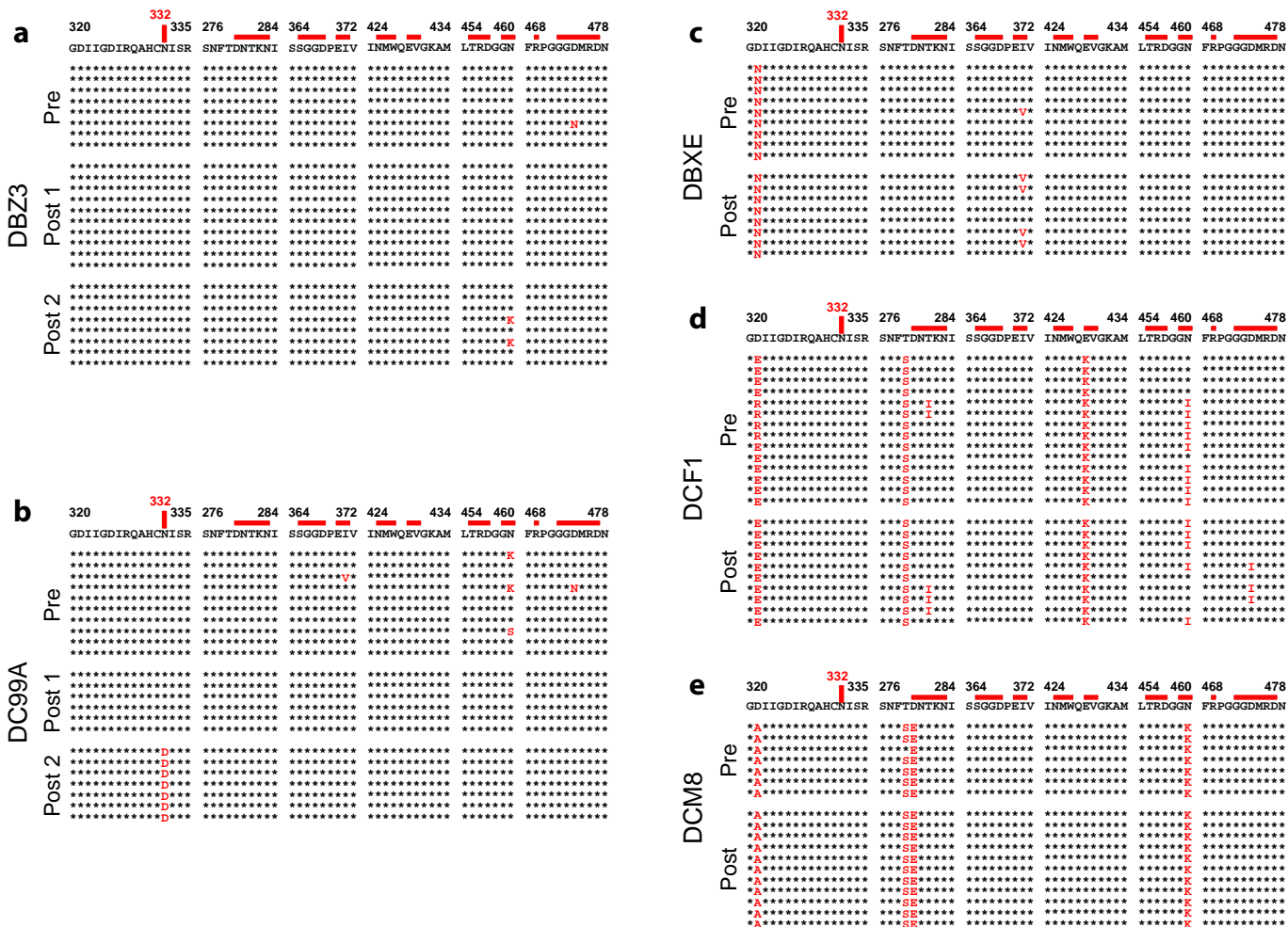
before (first 1,100 to 1,140 days) and during the first or second cycle of combination monoclonal antibody treatment are shown.





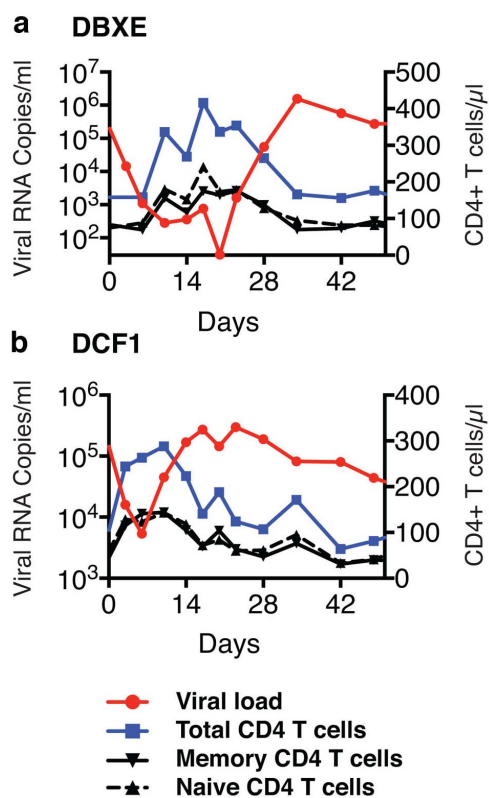
**Extended Data Figure 4** | BAL CD4<sup>+</sup> T cells in five chronically SHIV-infected macaques treated with two anti-HIV-1 neutralizing monoclonal antibodies. Plasma viral loads and the percentage of CD4<sup>+</sup> T cells in CD3<sup>+</sup>

gated BAL specimens, before (first 1,100 to 1,140 days) and during the first or second cycle of combination monoclonal antibody treatment, are shown.



**Extended Data Figure 5 | SGA analysis of selected SHIV-AD8EO gp120 sequences known to confer resistance to 10-1074 or 3BNC117 monoclonal antibody, before and after combination immunotherapy. a–e.** Plasmas from animals DBZ3 (pre ( $n = 8$ ); day 49 post first treatment ( $n = 10$ ); day 24 post second treatment ( $n = 8$ )) (a), DC99A (pre ( $n = 10$ ); day 57 post first treatment

( $n = 6$ ); day 41 post second treatment ( $n = 7$ )) (b), DBXE (pre ( $n = 9$ ), day 28 ( $n = 8$ )) (c), DCF1 (pre ( $n = 14$ ), day 28 ( $n = 10$ )) (d) and DCM8 (pre ( $n = 7$ ), day 28 ( $n = 11$ )) (e) were evaluated. The gp120 sequences at the top are present in the SHIV-AD8EO molecular clone inoculum. Mutations conferring resistance are highlighted in red.



**Extended Data Figure 6** | CD4<sup>+</sup> T-cell numbers increase during combination monoclonal antibody treatment of SHIV-AD8EO-infected macaques. **a, b**, Levels of viral RNA and total CD4<sup>+</sup> T-cell/CD4<sup>+</sup> T-cell subsets in symptomatic chronically infected macaques DBXE (**a**) and DCF1 (**b**).



**a**

	ID <sub>50</sub> (µg/ml)	
	10-1074	3BNC117
X2088_9	0.013	>100
Q769_d22	>100	0.024

**b**

	Plasma ID <sub>50</sub> Titer		Plasma ID <sub>50</sub> converted to respected antibody concentration of 10-1074 and 3BNC117 (µg/ml)	
Plasma + mAb	X2088_9	Q769_d22	X2088_9	Q769_d22
preP_0.01_3BNC117	<20	<20	<0.26	<0.48
preP_0.1_3BNC117	<20	<20	<0.26	<0.48
preP_1_3BNC117	<20	27	<0.26	0.65
preP_10_3BNC117	<20	340	<0.26	8.16
preP_100_3BNC117	<20	5,369	<0.26	128.86
preP_0.01_10-1074	<20	<20	<0.26	<0.48
preP_0.1_10-1074	<20	<20	<0.26	<0.48
preP_1_10-1074	54	<20	0.70	<0.48
preP_10_10-1074	516	<20	6.71	<0.48
preP_100_10-1074	11,419	<20	148.45	<0.48

preT - pre-treatment plasma of different NHPs (DBZ3, DC99A, DCM8, DCF1, DBXE)

ID <sub>50</sub> Titer	µg/ml
>5000	>75
>1000 - 5000	>15 - 75
>200 - 1000	>5 - 15
40 - 200	1 - 5

**Extended Data Figure 7 | Assays to identify 10-1074- or 3BNC117-specific neutralizing activities in the plasma of monoclonal-antibody-treated macaques.** **a**, ID<sub>50</sub> values measured in the TZM-bl neutralization assay of 10-1074 and 3BNC117 against HIV-1 strains that are sensitive to one but not the other broadly neutralizing antibody (that is, HIV-1 strain X2088\_9 (10-1074 sensitive); HIV-1 strain Q769\_d22 (3BNC117 sensitive)). **b**, Neutralizing

activities in plasma before antibody administration (preP), but spiked with 0.01, 0.1, 1, 10 and 100 µg ml<sup>-1</sup> of antibodies 10-1074 (blue) or 3BNC117 (green). Neutralizing activities are reported as plasma ID<sub>50</sub> titres (left columns) and converted to antibody concentrations (right columns) based on measured ID<sub>50</sub> values in **a**.

</

Extended Data Table 1 | Cell-associated viral RNA/DNA in rhesus macaques receiving combination monoclonal antibody therapy

Animal	Treatment Time (Days)	Plasma Viral RNA (copies/ml)	SIV Gag RNA Copies per 10 <sup>8</sup> Cell Eq	SIV Gag DNA Copies per 10 <sup>8</sup> Cell Eq
DBZ3 1st	0	1.08E+04	9,000	6,700
	10	< 100	360	7,500
	20	< 100	2,600	14,000
	24	< 100	1,600	6,400
	27	< 100	670	5,700
DBZ3 2nd	0	1.32E+04	52,000	15,000
	10	< 100	380	4,700
	13	4.09E+02	1,000	11,000
DC99A 1st	0	7.60E+03	31,000	1,400
	14	< 100	18,000	5,600
	20	< 100	8,100	2,700
	27	< 100	400	790
	34	< 100	550	1,100
	41	< 100	7,200	780
DC99A 2nd	0	2.35E+03	23,000	2,100
	21	< 100	570	1,100
	27	< 100	1,100	2,100
DBXE	0	1.96E+05	470,000	71,000
	14	3.59E+02	17,000	33,000
	17	7.55E+02	11,000	22,000
	20	< 100	11,000	33,000
	23	1.60E+03	17,000	27,000
DCM8	0	1.59E+03	110,000	8,600
	14	< 100	1,700	1,600
	17	< 100	880	5,000
	20	1.51E+02	22,000	6,600
DCF1	0	1.44E+05	240,000	15,000
	6	5.38E+03	34,000	5,500
	14	1.69E+05	190,000	11,000
	20	1.45E+05	1,100,000	14,000



**Extended Data Table 2 | Decay rate constants of SHIV-AD8 RNA in plasma after monoclonal antibody treatment**

<b>Animal</b>	<b>k (days<sup>-1</sup>)</b>	<b>T<sub>1/2</sub> (days)</b>
DC99A 1st	0.72	1.0
DC99A 2nd	0.45	1.5
DBZ3 1st	0.47	1.4
DBZ3 2nd	0.70	1.0
DBXE	0.66	1.0
DCF1	0.55	1.3
DCM8	0.46	1.5
KZ6	0.40	1.8
MB6	0.43	1.5
MB8	0.36	1.9
MCN	0.89	0.8

Rate constants of initial viral RNA decline in plasma (decay rate constants  $k$ ) and corresponding half-lives ( $t_{1/2}$ ) were determined as previously described (ref. 26). The decay rate constants  $k$  were calculated by fitting the data with an exponential function. The fit was very good with correlation coefficients in the range from 0.91 to 1.0.

# Inhibitory signalling to the Arp2/3 complex steers cell migration

Irene Dang<sup>1\*</sup>, Roman Gorelik<sup>1\*</sup>, Carla Sousa-Blin<sup>1\*</sup>, Emmanuel Derivery<sup>1</sup>, Christophe Guérin<sup>2</sup>, Joern Linkner<sup>3</sup>, Maria Nemethova<sup>4</sup>, Julien G. Dumortier<sup>5</sup>, Florence A. Giger<sup>5</sup>, Tamara A. Chipysheva<sup>6</sup>, Valeria D. Ermilova<sup>6</sup>, Sophie Vacher<sup>7</sup>, Valérie Campanacci<sup>8</sup>, Isaline Herrada<sup>9</sup>, Anne-Gaelle Planson<sup>8</sup>, Susan Fetics<sup>8</sup>, Véronique Henriot<sup>1</sup>, Violaine David<sup>1</sup>, Ksenia Oguievetskaia<sup>1</sup>, Goran Lakisic<sup>1</sup>, Fabienne Pierre<sup>1</sup>, Anika Steffen<sup>10</sup>, Adeline Boyreau<sup>11</sup>, Nadine Peyrières<sup>11</sup>, Klemens Rottner<sup>10,12</sup>, Sophie Zinn-Justin<sup>9</sup>, Jacqueline Cherfils<sup>8</sup>, Ivan Bièche<sup>7</sup>, Antonina Y. Alexandrova<sup>6</sup>, Nicolas B. David<sup>5</sup>, J. Victor Small<sup>4</sup>, Jan Faix<sup>3</sup>, Laurent Blanchoin<sup>2</sup> & Alexis Gautreau<sup>1</sup>

**Cell migration requires the generation of branched actin networks that power the protrusion of the plasma membrane in lamellipodia<sup>1,2</sup>. The actin-related proteins 2 and 3 (Arp2/3) complex is the molecular machine that nucleates these branched actin networks<sup>3</sup>. This machine is activated at the leading edge of migrating cells by Wiskott–Aldrich syndrome protein (WASP)-family verprolin-homologous protein (WAVE, also known as SCAR). The WAVE complex is itself directly activated by the small GTPase Rac, which induces lamellipodia<sup>4–6</sup>. However, how cells regulate the directionality of migration is poorly understood. Here we identify a new protein, Arpin, that inhibits the Arp2/3 complex *in vitro*, and show that Rac signalling recruits and activates Arpin at the lamellipodial tip, like WAVE. Consistently, after depletion of the inhibitory Arpin, lamellipodia protrude faster and cells migrate faster. A major role of this inhibitory circuit, however, is to control directional persistence of migration. Indeed, Arpin depletion in both mammalian cells and *Dictyostelium discoideum* amoeba resulted in straighter trajectories, whereas Arpin micro-injection in fish keratocytes, one of the most persistent systems of cell migration, induced these cells to turn. The coexistence of the Rac–Arpin–Arp2/3 inhibitory circuit with the Rac–WAVE–Arp2/3 activatory circuit can account for this conserved role of Arpin in steering cell migration.**

The Arp2/3 complex is activated at different cellular locations by different nucleation promoting factors (NPFs), WAVE at lamellipodia, neural WASP (N-WASP) at clathrin-coated pits, and WASP and SCAR homologue (WASH) at endosomes<sup>7,8</sup>. NPFs share a characteristic carboxy-terminal tripartite domain, referred to as the VCA<sup>9</sup>. The A (acidic) motif binds to the Arp2/3 complex and induces its conformational activation. Arp2/3 inhibitory proteins containing an A motif, PICK1 and Gadkin (also known as AP1AR and  $\gamma$ -BAR), were detected at endocytic pits and at endosomes<sup>10,11</sup>. Thus, although endocytic pits and endosomes possess antagonistic activities towards the Arp2/3 complex, it was not known whether lamellipodia contain a similar Arp2/3 inhibitory protein to counteract WAVE. To identify such a protein, we performed a bioinformatics search for proteins displaying the typical A motif of human NPFs, characterized by a tryptophan residue at the antepenultimate position in an acidic context. We identified an uncharacterized protein (C15orf38) clustered with NPFs with this procedure (Fig. 1a). This protein was named Arpin. Arpin is encoded by a single gene in metazoans and amoeba. Predictions and NMR analysis indicate that this protein of about 220 residues is structured

with the exception of its highly mobile C-terminal end, which contains the putative Arp2/3-binding site (Extended Data Figs 1 and 2). Indeed, Arpin binds to Arp2/3 mostly through its acidic motif (Fig. 1b and Extended Data Figs 3 and 4).

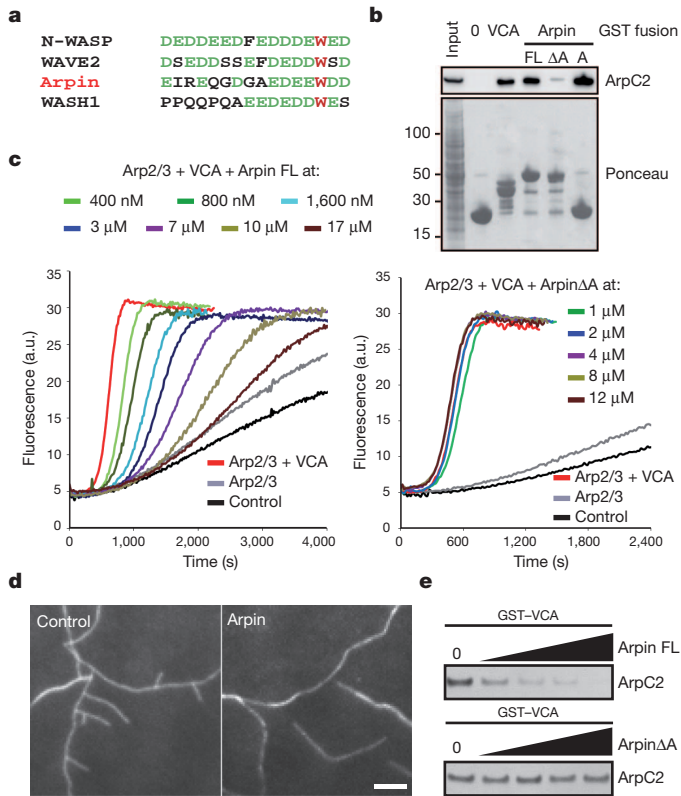
The molecular function of Arpin on Arp2/3 activity was assayed by spectrofluorimetry and total internal reflection fluorescence (TIRF) microscopy using purified proteins. Arpin was unable to activate the Arp2/3 complex, consistent with its lack of verprolin (V) and cofilin (C) homology motifs. However, when Arp2/3 was activated by VCA, Arpin, but not its truncated form lacking the acidic motif (Arpin $\Delta$ A), inhibited actin polymerization in a dose-dependent manner (Fig. 1c). The acidic peptide was sufficient for this inhibition, although it was less effective than full-length Arpin, in line with its lower affinity for the Arp2/3 complex (Extended Data Fig. 4). Arpin inhibited Arp2/3 activation, because we observed by TIRF microscopy the generation of fewer actin branched junctions in the presence of Arpin (Fig. 1d and Supplementary Video 1). Full-length Arpin and its acidic motif, but not Arpin $\Delta$ A, compete with VCA for Arp2/3 binding (Fig. 1e and Extended Data Fig. 4). Therefore, Arpin is a new competitive inhibitor of the Arp2/3 complex. The name Arpin is a mnemonic for its activity (Arp2/3 inhibition).

The subcellular localization of Arpin was examined by immunofluorescence in spreading mouse embryonic fibroblasts (MEFs). Arpin was detected in restricted segments of the plasma membrane, which were also stained by three lamellipodial markers—the WAVE complex, the Arp2/3 complex and cortactin (Fig. 2a and Extended Data Fig. 5). Radial line scans of immunofluorescence pictures through lamellipodial outlines were generated, registered with the edge as a reference, and averaged to reveal the relative distributions of these different lamellipodial components. Arpin overlapped perfectly with the distribution of the WAVE complex, a tip component<sup>12</sup> (Fig. 2a). Arpin is thus localized at the lamellipodium tip, where new actin branches are nucleated by WAVE and Arp2/3 complexes. As expected, Arp2/3 and cortactin distribution extended rearwards compared to Arpin (Extended Data Fig. 5), because Arp2/3 and cortactin correspond to branches of the lamellipodial actin network undergoing retrograde flow with respect to the protruding membrane<sup>9,13</sup>.

To understand the regulation of Arpin activity, we examined the role of Rac, the master controller of lamellipodium formation. We co-transfected 293T cells with different forms of Rac and green fluorescent protein (GFP)–Arpin and then analysed the interaction of

<sup>1</sup>Group Cytoskeleton in Cell Morphogenesis, Laboratoire d'Enzymologie et Biochimie Structurales, CNRS UPR3082, Gif-sur-Yvette 91190, France. <sup>2</sup>Institut de Recherches en Technologies et Sciences pour le Vivant (iRTSV), Laboratoire de Physiologie Cellulaire et Végétale, CNRS/CEA/INRA/UJF, Grenoble 38054, France. <sup>3</sup>Institute for Biophysical Chemistry, Hannover Medical School, Hannover 30625, Germany. <sup>4</sup>Institute of Molecular Biotechnology, Vienna 1030, Austria. <sup>5</sup>INSERM U1024, CNRS UMR8197, ENS, Institut de Biologie de l'ENS, Paris 75005, France. <sup>6</sup>Institute of Carcinogenesis, N. N. Blokhin Cancer Research Center, Russian Academy of Medical Sciences, Moscow 115478, Russia. <sup>7</sup>Oncogenetic Laboratory, Institut Curie, Hôpital René Huguenin, Saint-Cloud 92210, France. <sup>8</sup>Group Small G Proteins, Laboratoire d'Enzymologie et Biochimie Structurales, CNRS UPR3082, Gif-sur-Yvette 91190, France. <sup>9</sup>Laboratoire de Biologie Structurale et Radiobiologie (iBiTec-S), CNRS URA2096, CEA Saclay, Gif-sur-Yvette 91190, France. <sup>10</sup>Institute of Genetics, University of Bonn, Bonn 53115, Germany. <sup>11</sup>Institut des Systèmes Complexes & NeD, Institut de Neurobiologie Alfred Fessard, CNRS UPR3294, Gif-sur-Yvette 91190, France. <sup>12</sup>Helmholtz Centre for Infection Research, Braunschweig 38124, Germany.

\*These authors contributed equally to this work.

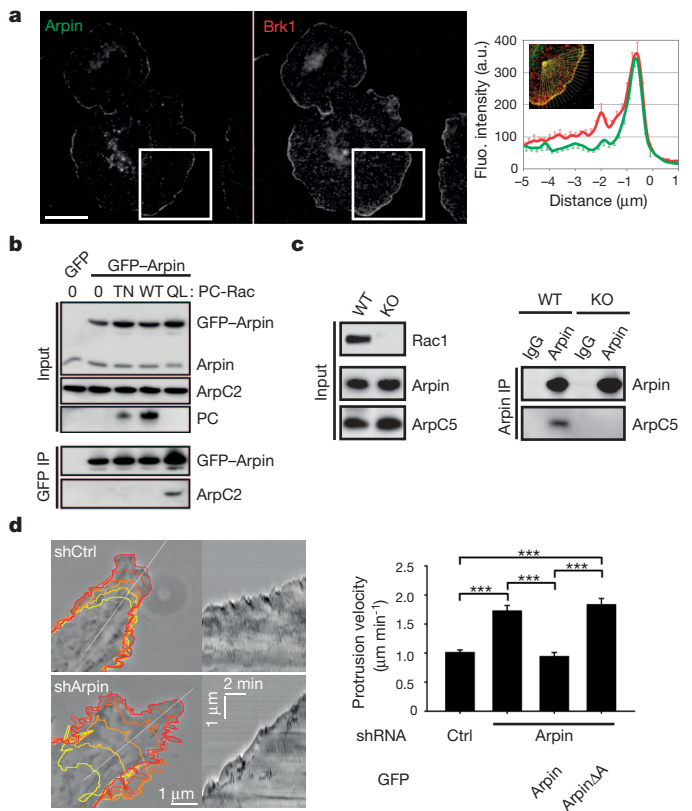


**Figure 1 | Arpin inhibits Arp2/3 activation *in vitro*.** **a**, Alignment of acidic C termini of three NPFs and of Arpin. **b**, Arpin binds to the Arp2/3 complex through its acidic C-terminal region. Glutathione S-transferase (GST) pull-down with full-length Arpin (FL), its last 16 amino-acids (A), ArpinΔA or the VCA domain of N-WASP as a positive control. ArpC2 is a subunit of the Arp2/3 complex. **c**, Spectrofluorimetry assay using pyrene-labelled actin to monitor polymerization. a.u., arbitrary units. **d**, Assembly of branched actin networks monitored by TIRF microscopy using rhodamine-labelled actin. Scale bar, 5 μm. **e**, Full-length Arpin, but not ArpinΔA (18 μM and serial twofold dilutions), competes with the NPF for Arp2/3 binding.

Arpin with the Arp2/3 complex through GFP immunoprecipitations. The active form of Rac1, which is sufficient to induce lamellipodia, was also sufficient to induce Arp2/3 co-immunoprecipitation with Arpin (Fig. 2b). To examine whether Rac is required for Arpin activation, we used Rac1 knockout MEFs that lack lamellipodia<sup>14</sup>. The absence of Rac abrogated the peripheral localization of Arpin in all knockout MEF cells examined (Extended Data Fig. 5). Endogenous Arpin was then immunoprecipitated from Rac1 knockout MEFs. The Arp2/3 complex co-immunoprecipitated with Arpin in control MEF cells, but not in Rac-deficient cells (Fig. 2c). Together, these results show that, in response to Rac signalling, Arpin inhibits the Arp2/3 complex at the lamellipodium tip, that is, where Rac also stimulates actin polymerization through the WAVE complex.

This counterintuitive finding suggests that Arpin would be a built-in brake of protrusions. Human RPE1 cells transiently transfected with short hairpin RNAs (shRNA) that efficiently deplete Arpin displayed increased lamellipodia-mediated cell spreading (Extended Data Fig. 6 and Supplementary Video 2). Arpin depletion increased protrusion velocity of lamellipodia, consistent with its Arp2/3 inhibitory role (Fig. 2d). This effect was fully rescued by Arpin re-expression, but only when Arpin contained the acidic motif. Arpin thus provides a paradoxical negative circuit downstream of Rac. Such a circuitry, where Rac induces and inhibits actin polymerization, generates an 'incoherent feedforward loop' (Fig. 2e), which can favour temporal regulations<sup>15</sup>.

To examine whether the Arpin circuit is physiologically relevant for cell migration, we impaired the expression of the *arpin* gene in zebrafish embryos using morpholinos. During gastrulation, prechordal plate



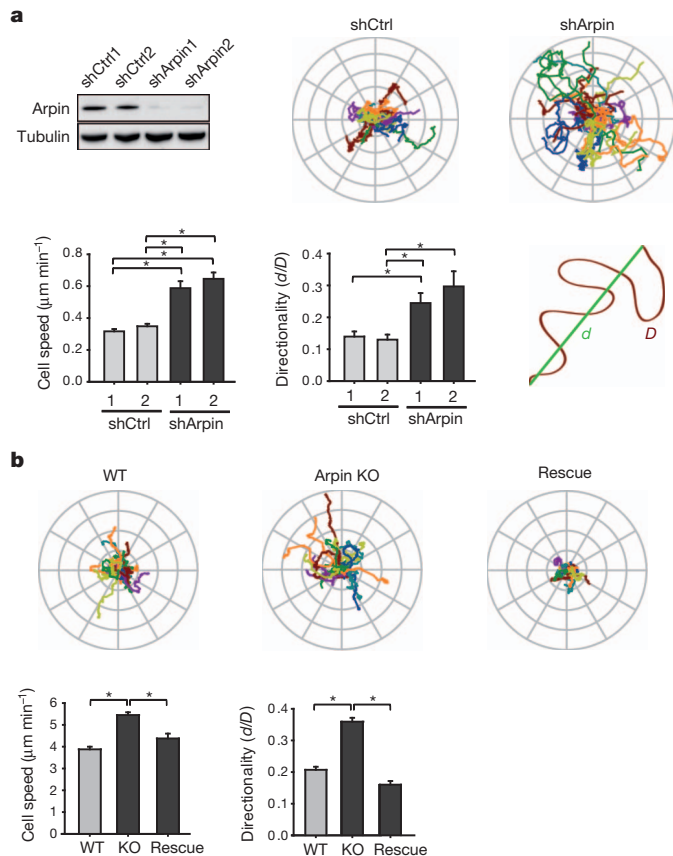
**Figure 2 | Arpin inhibits the Arp2/3 complex at the lamellipodium tip.**

**a**, Arpin colocalizes with Brk1, a subunit of the WAVE complex, at the lamellipodium tip (mean ± s.e.m., 24 line scans). Scale bar, 20 μm. **b**, Active (Gln61Leu; QL), but not wild-type (WT), nor dominant negative (Thr17Asn; TN), Rac induces Arpin association with the Arp2/3 complex. IP, immunoprecipitation. PC, protein C epitope. **c**, Rac is required for the Arpin-Arp2/3 interaction. KO, knockout. **d**, Arpin depletion increases the speed at which lamellipodia protrude. GFP-Arpin expression rescues the phenotype. Ctrl, control; shCtrl, control shRNA; shArpin, *arpin* shRNA. Data are mean ± s.e.m.; *n* = 40; \*\*\**P* < 0.001, two-tailed analysis of variance (ANOVA). **e**, Arpin circuitry.

cells collectively migrate towards the animal pole. After *arpin* loss of function, cell movements were less coordinated (Extended Data Fig. 7). Prechordal plate cell transplants revealed a cell autonomous effect of Arpin on protrusions. Protrusions were more frequent and more persistent over time in the absence of Arpin (Supplementary Video 3). This observation is consistent with the incoherent feedforward loop, a circuitry that can suppress the protrusion it creates.

To understand the role of the incoherent feedforward loop in cell migration further, we performed Arpin loss-of-function experiments in cell systems migrating as individual cells. Migration of stable Arpin-depleted clones from the breast human cell line MDA-MB-231 was analysed by video microscopy in two or three dimensions (Fig. 3a, Supplementary Videos 4 and 5, and Extended Data Fig. 8). In both cases, the tracks illustrate that Arpin-depleted cells explored a larger territory than control cells, an observation substantiated by mean square displacements (Extended Data Fig. 9). Increased exploration was not only due to increased speed, but also to increased directional persistence, measured as the ratio of the distance between two points by the actual trajectory (Fig. 3a) or as the direction autocorrelation function (Extended Data Fig. 10). Because Arpin is conserved in amoeba,



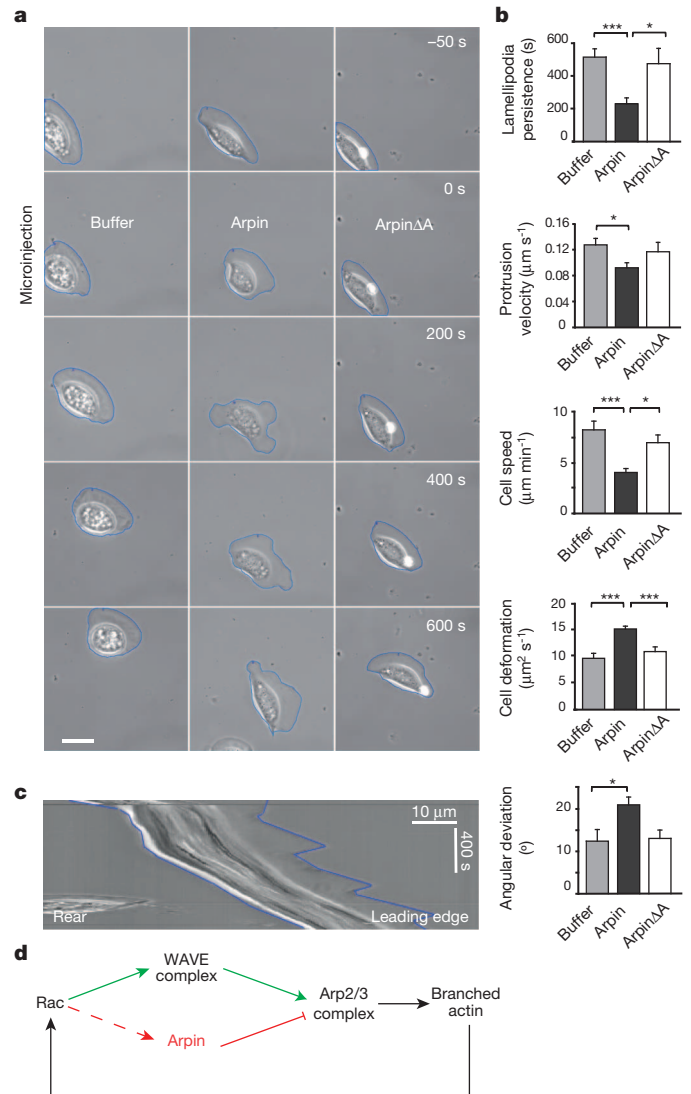


**Figure 3 | Arpin depletion increases directional persistence of migration in mammalian cells and in *Dictyostelium discoideum*.** **a**, Arpin-depleted (shArpin1 and shArpin2) MDA-MB-231 cells explore a wider territory than controls owing to increased speed and directional persistence. The directionality index is calculated as the ratio of the distance between starting and ending points ( $d$ ) by the actual trajectory ( $D$ ). Mean  $\pm$  s.e.m.;  $n = 24, 31, 32$  and  $26$ , respectively;  $*P < 0.05$ , two-tailed Kruskal–Wallis test. **b**, Arpin-knockout amoeba explores a wider territory than wild type owing to increased speed and directional persistence. Directionality is more than fully rescued by *Dictyostelium* Arpin expression in knockout amoeba. Data are mean  $\pm$  s.e.m.;  $n = 116, 196$  and  $45$ , respectively;  $*P < 0.05$ , two-tailed Kruskal–Wallis test.

we analysed its function in *Dictyostelium discoideum* by knocking out the orthologous gene. As in mammalian cells, Arpin-knockout *Dictyostelium* amoebae explored a wider territory than the controls owing to increased cell speed and directional persistence (Fig. 3b and Supplementary Video 6). However, directional persistence, which is more than fully rescued by GFP–Arpin expression, is the most affected parameter in amoeba.

Arpin could thus be a ‘steering factor’. To test this hypothesis directly in a gain-of-function experiment, we selected the fish keratocyte model. These cells are characterized by fast migration based on a wide fan-shaped lamellipodium with high directional persistence. We purified zebrafish Arpin and Arpin $\Delta A$  from *Escherichia coli* and microinjected these proteins into migrating trout keratocytes (Fig. 4a). Injection of Arpin, but not Arpin $\Delta A$ , caused keratocytes to reduce their speed and deviate from their initial direction of migration (Fig. 4b and Extended Data Fig. 10). Notably, Arpin did not prevent lamellipodia protrusion, but resulted in cycles of suppression of existing lamellipodia followed by formation of new, ectopic ones (Fig. 4c and Supplementary Video 7). Collectively, the experiments performed in different systems of cell migration support a role for Arpin in promoting cell steering: Arpin slows cells down and allows them to turn.

In a computational model of efficient and persistent cell migration, the lamellipodium spatially determines where the WAVE and the Arp2/3 complexes will next polymerize actin, thus maintaining the



**Figure 4 | Arpin microinjection induces fish keratocyte to turn.** **a**, Gallery of fish keratocytes microinjected with purified full-length Arpin, Arpin $\Delta A$  or buffer as control. Scale bar, 20  $\mu\text{m}$ . **b**, Quantifications. Data are mean  $\pm$  s.e.m.;  $n = 16, 15$  and  $11$ , respectively;  $*P < 0.05$ ,  $***P < 0.001$ , two-tailed ANOVA. **c**, Kymograph of the Arpin-microinjected keratocyte. **d**, Model.

front at the front over time<sup>16</sup> (Fig. 4d). Such a feedback loop, sensing where branched actin is polymerized and activating Rac as a response, has recently been identified: it involves coronin 1A and the Rac exchange factor  $\beta$ -Pix<sup>17</sup>. In this feedback system, the WAVE complex closes a positive feedback loop that maintains efficient directional migration over time, whereas Arpin closes a concurrent negative feedback loop, which induces braking and allows turning. These two nested feedback loops, positive and negative, can account for the emergence of oscillations in lamellipodium protrusion/retraction, as observed in fish keratocytes after Arpin injection (Supplementary Video 8), and for various travelling actin waves described in different systems<sup>18–21</sup>.

Other proteins were previously shown to regulate cell steering<sup>22,23</sup>. Knockdown of Rac1 or cofilin, a protein that severs and depolymerizes actin filaments, increases directional persistence of mammalian cells<sup>24,25</sup>. These proteins are required, however, for lamellipodial protrusion and actin-based motility<sup>26,27</sup>. Arpin is unique in that it regulates cell steering, while being dispensable for lamellipodial protrusion and efficient migration. Arpin is a prime candidate to fine-tune numerous physiological migrations biased by diverse cues<sup>23</sup>. Arpin also seems to have a role in preventing cells from migrating. In this respect, dissection of the

mechanisms regulating Arpin in physiology and pathology is a major challenge ahead of us.

## METHODS SUMMARY

**Arpin purification and antibody production.** Arpin was purified from *E. coli* as a glutathione S-transferase (GST) fusion protein and cleaved off GST. Untagged Arpin was used for *in vitro* assays of actin polymerization, antibody generation and keratocyte injection. Rabbit polyclonal antibodies targeting full-length Arpin were purified by affinity chromatography. Immunoprecipitation of endogenous Arpin was performed using purified antibodies coupled to magnetic beads.

**Cells and imaging.** RPE1 cells were electroporated with plasmids. Plasmids encoding fusion proteins or shRNAs were transfected into MDA-MB-231 cells using liposomes. Lamellipodial dynamics and random migration were analysed with ImageJ using the plug-ins 'Kymograph' and 'MtrackJ', respectively. Live-cell imaging was performed using an inverted Axio Observer microscope (Zeiss) equipped with two chambers controlled for temperature and CO<sub>2</sub>. For Arpin localization, MEF cells were fixed with 10% TCA, permeabilized with 0.2% Triton X-100, and processed for immunofluorescence. To draw radial line scans, a custom-made ImageJ plug-in was developed, edge was determined using 'Isodata' and 'Analyze particle', and then a custom-made VBA macro in Excel was used to align data relative to the edge. Keratocytes were isolated from scales of freshly killed brook trout (*Salvelinus fontinalis*) and microinjected with a Femtojet (Eppendorf). Plots were drawn with SigmaPlot (SPSS) and statistics were performed with SigmaStat (SPSS).

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 November 2012; accepted 28 August 2013.

Published online 16 October 2013.

- Wu, C. *et al.* Arp2/3 is critical for lamellipodia and response to extracellular matrix cues but is dispensable for chemotaxis. *Cell* **148**, 973–987 (2012).
- Suraneni, P. *et al.* The Arp2/3 complex is required for lamellipodia extension and directional fibroblast cell migration. *J. Cell Biol.* **197**, 239–251 (2012).
- Mullins, R. D., Heuser, J. A. & Pollard, T. D. The interaction of Arp2/3 complex with actin: nucleation, high affinity pointed end capping, and formation of branching networks of filaments. *Proc. Natl Acad. Sci. USA* **95**, 6181–6186 (1998).
- Insall, R. H. & Machesky, L. M. Actin dynamics at the leading edge: from simple machinery to complex networks. *Dev. Cell* **17**, 310–322 (2009).
- Padrick, S. B. & Rosen, M. K. Physical mechanisms of signal integration by WASP family proteins. *Annu. Rev. Biochem.* **79**, 707–735 (2010).
- Ridley, A. J. Life at the leading edge. *Cell* **145**, 1012–1022 (2011).
- Derivery, E. *et al.* The Arp2/3 activator WASH controls the fission of endosomes through a large multiprotein complex. *Dev. Cell* **17**, 712–723 (2009).
- Suetsugu, S. & Gautreau, A. Synergistic BAR-NPF interactions in actin-driven membrane remodeling. *Trends Cell Biol.* **22**, 141–150 (2012).
- Pollard, T. D. Regulation of actin filament assembly by Arp2/3 complex and formins. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 451–477 (2007).
- Rocca, D. L., Martin, S., Jenkins, E. L. & Hanley, J. G. Inhibition of Arp2/3-mediated actin polymerization by PICK1 regulates neuronal morphology and AMPA receptor endocytosis. *Nature Cell Biol.* **10**, 259–271 (2008).
- Maritzen, T. *et al.* Gadin negatively regulates cell spreading and motility via sequestration of the actin-nucleating ARP2/3 complex. *Proc. Natl Acad. Sci. USA* **109**, 10382–10387 (2012).
- Lai, F. P. *et al.* Arp2/3 complex interactions and actin network turnover in lamellipodia. *EMBO J.* **27**, 982–992 (2008).
- Cai, L., Makhov, A. M., Schafer, D. A. & Bear, J. E. Coronin 1B antagonizes cortactin and remodels Arp2/3-containing actin branches in lamellipodia. *Cell* **134**, 828–842 (2008).
- Steffen, A. *et al.* Rac function is critical for cell migration but not required for spreading and focal adhesion formation. *J. Cell Sci.* <http://dx.doi.org/10.1242/jcs.118232> (31 July 2013).
- Hart, Y. & Alon, U. The utility of paradoxical components in biological circuits. *Mol. Cell* **49**, 213–221 (2013).
- Neilson, M. P. *et al.* Chemotaxis: a feedback-based computational model robustly predicts multiple aspects of real cell behaviour. *PLoS Biol.* **9**, e1000618 (2011).
- Castro-Castro, A. *et al.* Coronin 1A promotes a cytoskeletal-based feedback loop that facilitates Rac1 translocation and activation. *EMBO J.* **30**, 3913–3927 (2011).
- Machacek, M. & Danuser, G. Morphodynamic profiling of protrusion phenotypes. *Biophys. J.* **90**, 1439–1452 (2006).
- Weiner, O. D., Marganski, W. A., Wu, L. F., Altschuler, S. J. & Kirschner, M. W. An actin-based wave generator organizes cell motility. *PLoS Biol.* **5**, e221 (2007).
- Brandman, O. & Meyer, T. Feedback loops shape cellular signals in space and time. *Science* **322**, 390–395 (2008).
- Allard, J. & Mogilner, A. Traveling waves in actin dynamics and cell motility. *Curr. Opin. Cell Biol.* **25**, 107–115 (2013).
- Ghosh, M. *et al.* Cofilin promotes actin polymerization and defines the direction of cell motility. *Science* **304**, 743–746 (2004).
- Petrie, R. J., Doyle, A. D. & Yamada, K. M. Random versus directionally persistent cell migration. *Nature Rev. Mol. Cell Biol.* **10**, 538–549 (2009).
- Pankov, R. *et al.* A Rac switch regulates random versus directionally persistent cell migration. *J. Cell Biol.* **170**, 793–802 (2005).
- Sidani, M. *et al.* Cofilin determines the migration behavior and turning frequency of metastatic cancer cells. *J. Cell Biol.* **179**, 777–791 (2007).
- Li, A. *et al.* Rac1 drives melanoblast organization during mouse development by orchestrating pseudopod-driven motility and cell-cycle progression. *Dev. Cell* **21**, 722–734 (2011).
- Loisel, T. P., Boujema, R., Pantaloni, D. & Carlier, M. F. Reconstitution of actin-based motility of *Listeria* and *Shigella* using pure proteins. *Nature* **401**, 613–616 (1999).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** A.G. acknowledges his PhD supervisor M. Arpin, the name of the here identified protein is a tribute to her mentoring. We thank G. Romet-Lemonne, E. Portnoy and F. Marletaz for suggestions. We acknowledge support from Agence Nationale pour la Recherche (ANR-08-BLAN-0012-CSD 8 to A.G. and L.B., ANR-08-PCVI-0010-03 to A.G., ANR-11-BSV8-0010-02 to A.G., J.C. and S.Z.-J.), Association pour la Recherche sur le Cancer (SFI20101201512 to A.G., PDF20111204331 to R.G., SFI20111203770 to N.B.D.), the Bio-Emergences IBISA facility and Fundacao para a Ciencia e a Tecnologia (SFRH/BPD/46451/2008 to C.S.-B.), the Austrian Science Fund (FWF P21292-B09 to J.V.S.), the Deutsche Forschungsgemeinschaft (FA 330/5-1 to J.F.) and grant number 8066, code 2012-1.1-12-000-1002-064 from the Russian Ministry of Education and Science to A.Y.A.

**Author Contributions** I.D., R.G. and C.S.-B. performed videomicroscopy, analysed cell migration, analysed biochemical interactions of Arpin and its localization. E.D. wrote the bioinformatics programme that first identified Arpin. C.G. and L.B. performed *in vitro* actin polymerization and fluorescence anisotropy assays. J.L. and J.F. isolated knockout amoeba and analysed their migration. M.N. and J.V.S. micro-injected fish keratocytes. J.G.D., F.A.G. and N.B.D. characterized the Arpin phenotype in zebrafish. A.B. and N.P. determined the Arpin expression profile in zebrafish. I.H. and S.Z.-J. contributed the NMR spectrum. T.A.C., V.D.E., A.Y.A., S.V., I.B., V.C., V.D., G.L., K.O., F.P., A.-G.P., S.F. and V.H. generated DNA constructs, isolated stable cell clones, purified and characterized recombinant proteins, and performed crucial experiments for our understanding of Arpin function. A.S. and K.R. isolated the Rac1 knockout MEFs. All authors designed experiments. N.P., K.R., S.Z.-J., J.C., N.B.D., I.B., A.Y.A., J.V.S., J.F., L.B. and A.G. supervised the work in their respective research group. A.G. coordinated the study and wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.G. ([alexis.gautreau@lebs.cnrs-gif.fr](mailto:alexis.gautreau@lebs.cnrs-gif.fr))

## METHODS

**Plasmids.** Human and zebrafish Arpin were amplified by PCR from clones IMAGE:5770387 and IMAGE:7404342, respectively (GeneService). *Dictyostelium discoideum* DdArpin was amplified from Ax2 cDNA. Human full-length Arpin (residues 1–226), ArpinΔA (residues 1–210) or ArpinA (residues 211–226), zebrafish full-length Arpin (residues 1–226), ArpinΔA (residues 1–210), and murine N-WASP VCA fragment (residues 392–501)<sup>28</sup> were cloned into a modified pGEX vector with a TEV cleavage site between the restriction sites FseI and AscI. For expression in mammalian cells, Arpin inserts were cloned into a compatible plasmid pcDNAm PC-GFP blue<sup>7</sup>. Zebrafish full-length Arpin was also inserted pBluescript to generate probes for *in situ* hybridization and in pCS2-GFP for rescue experiments. Human RAC1 wild-type, Thr17Asn, Gln61Leu, the Arp2/3 complex subunits ArpC5A and ArpC5B<sup>29</sup> were cloned into pcDNA5 His PC TEV blue<sup>7</sup>. For expression in amoeba, *Dictyostelium* Arpin was inserted into pDGFP-MCS-neo<sup>30</sup>. For shRNA-expressing plasmids, two hybridized oligonucleotides (MWG) were cloned into psiRNA-h7SKblast1 G1 (Invivogen) according to the manufacturer's protocol. The following target sequences were used: shArpin 1: 5'-GGAGAACTGATCGATGTATCT-3'; shArpin 2: 5'-GCTTCCTCATGT CGTCTACA-3'; shArpin 3: 5'-GCCTTCCTAGACATTACATGA-3' (targets the 3' untranslated region (UTR) of *arpin* mRNA); shArpC2 1: 5'-CATGTATGT TGAGTCTAA-3'; and shArpC2 2: 5'-GCTCTAAGGCCTATATTCA-3'. These plasmids were compared to the non-targeting control provided by Invivogen; shControl: 5'-GCATATGTGCGTACCTAGCAT-3'. All constructs were verified by sequencing.

**Protein purification.** Arpin, ArpinΔA, ArpinA and N-WASP VCA fused to GST were purified from *E. coli* BL21\* strain (Life Technologies) using standard purification protocols, dialysed against storage buffer (20 mM Tris-HCl, 50 mM NaCl, 1 mM dithiothreitol (DTT), pH 7.5), frozen in liquid nitrogen and stored at -80 °C. When indicated, Arpin was cleaved by TEV protease off GST. Arpin bound to glutathione sepharose 4B beads was cleaved by overnight incubation at 4 °C using His-tagged TEV protease in 50 mM Tris, pH 7.5, 2 mM β-mercaptoethanol, 100 mM NaCl and 5 mM MgCl<sub>2</sub>. TEV was removed by incubation with Ni<sup>2+</sup> beads (GE Healthcare). Arpin was further purified by size-exclusion chromatography (SEC) on a Superdex-200 column (GE Healthcare) and concentrated on Vivaspins filters. Human Arpin was used for the production of polyclonal antibodies and competition experiments. Zebrafish Arpin was similarly produced, purified and used for keratocyte injection at 7.5 μg μl<sup>-1</sup> in 15 mM Tris-HCl, 150 mM NaCl, 5 mM MgCl<sub>2</sub> and 1 mM DTT, pH 7.5. Both proteins had an amino-terminal extension of 10 amino acids (GAMAHMGRP) after TEV cleavage. ArpinA peptide (residues 211–226 of full-length Arpin) was purchased from Proteogenix. For the SEC coupled to multiangle light scattering (SEC-MALS) characterization, proteins were separated in a 15-ml KW-803 column (Shodex) run on a Shimadzu HPLC system. MALS, quasi-elastic light scattering (QELS) and refractive index (RI) measurements were achieved with a MiniDawn Treos (Wyatt technology), a WyattQELS (Wyatt technology) and an Optilab T-reX (Wyatt technology), respectively. Molecular mass and hydrodynamic radius calculations were performed with the ASTRA VI software (Wyatt Technology) using a *dn/dc* value of 0.183 ml g<sup>-1</sup>.

**Antibodies.** Polyclonal antibodies targeting Arpin were obtained in rabbits (AgroBio) against purified human Arpin, and then purified by affinity purification on a HiTrap NHS-activated HP column (GE Healthcare) coupled to the immunogen. ArpC2 polyclonal antibody and cortactin monoclonal antibody (clone 4F11) were from Millipore. ArpC5 monoclonal antibody (clone 323H3) was from Synaptic Systems. Brk1 monoclonal antibody (clone 231H9) was described earlier<sup>31</sup>. Tubulin monoclonal antibody (clone E7) was obtained from Developmental Studies Hybridoma Bank. PC monoclonal antibody (clone HPC4) was from Roche.

**In vitro assays of actin polymerization.** Pyrene actin assays and monitoring of the branching reaction were performed as described previously<sup>32</sup>. VCA refers to the VCA domain of WAVE1 purified as described<sup>33</sup>. Conditions for Fig. 1c were: 2 μM actin (10% pyrene-labelled), 500 nM VCA, 20 nM Arp2/3 and Arpin full-length or ArpinΔA at the indicated concentrations. Conditions for Fig. 1d were: 1 μM actin (10% rhodamine-labelled), 150 nM VCA, 80 nM Arp2/3 and 5 μM Arpin when indicated.

**Fluorescence anisotropy based determination of K<sub>d</sub>.** The ArpinA peptide was synthesized and labelled with 5-TAMRA at the amino terminus (Proteogenix). The peptide was excited with polarized light at 549 nm and emitted light was detected at 573 nm using a MOS450 fluorimeter (Biologic). Measurements were made for 60 s at 1 point s<sup>-1</sup>, and the average anisotropy was calculated with the Biologic software. Fits were performed as described previously<sup>34</sup>.

**GST pull-down, immunoprecipitations, SDS-PAGE and western blots.** HeLa cells were lysed in 50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.5% Triton X-100 and 5% glycerol, pH 7.5. GST fusion protein (20 μg) associated with 20 μl of glutathione sepharose 4B beads (GE Healthcare) was incubated with

1 ml HeLa cell extracts for 2 h at 4 °C. Beads were washed and analysed by western blot.

Co-immunoprecipitation of Arpin with the Arp2/3 complex was performed with either two 15-cm dishes of MEF cells, or one 10-cm dish of transfected 293T cells. Cell lysates prepared in 10 mM HEPES, pH 7.7, 50 mM KCl, 1 mM MgCl<sub>2</sub>, 1 mM EGTA and 1% Triton X-100 were incubated with 10 μg of non-immune rabbit IgG or 10 μg of anti-Arpin antibodies coupled to tosyl-activated dynabeads (Life Technologies) or to GFP-trap beads (Chromotek). Beads were incubated with extracts for 2 h at 4 °C, washed and analysed by western blot.

SDS-PAGE was performed using NuPAGE 4–12% Bis-Tris gels (Life Technologies). For western blots, nitrocellulose membranes were developed using horseradish peroxidase (HRP)-coupled antibodies, Supersignal kit (Pierce) and a LAS-3000 imager (Fujifilm).

**Cells and transfections.** hTERT immortalized RPE1 cells (Clontech) were grown in DMEM/HAM F12, MEFs and 293T cells in DMEM, and MDA-MB-231 cells were grown in RPMI. All media were supplemented with 10% FBS (media and serum from PAA Laboratories). All cells and stable clones were found negative for mycoplasma infection by a sensitive PCR assay.

RPE1 cells were electroporated with ECM 630 BTX (Harvard Apparatus). Ten-million cells were resuspended in 200 μl serum-free DMEM/HAM F12 medium containing 7.5 mM HEPES, pH 7.5, mixed with 10–40 μg DNA plasmid in 50 μl 210 mM NaCl and electroporated at 1,500 μF and 250 V. To isolate stable Arpin-depleted clones, MDA-MB-231 cells were transfected with shRNA Arpin 3 or shControl using Lipofectamine 2000, and clones selected with 10 μg ml<sup>-1</sup> blasticidin (Invivogen) were isolated with cloning rings and expanded. For rescue experiments, cells transfected with shRNA 3, which targets the 3' UTR, were transfected with GFP-Arpin, which lacks UTR sequences. To validate Arpin localization, MEFs were transfected with non-targeting (D-001810-10) or Arpin targeting (J-059240-10; ON-TARGET plus siRNA, Dharmacon) using lipofectamine RNAiMax (Life Technologies), and examined after 2 days.

**Immunofluorescence and live imaging of mammalian cells.** Cells were fixed in 10% TCA, permeabilized with 0.2% Triton X-100, and processed for immunofluorescence. To draw radial line scans, a custom made ImageJ plug-in was developed, edge was determined using 'Isodata' thresholding, then a custom-made VBA macro in Excel was used to align data relative to the edge. Lamellipodial dynamics and random migration were analysed with ImageJ using the plugins 'Kymograph' and 'MtrackJ', respectively. All imaging was done on an Axio Observer microscope (Zeiss) equipped with a Plan-Apochromat 63×/1.40 oil immersion objective, an EC Plan-Neofluar 40×/1.30 oil immersion objective and a Plan-Apochromat 20×/0.80 air objective, a Hamamatsu camera C10600 Orca-R<sup>2</sup> and a Pecon Zeiss incubator XL multi S1 RED LS (Heating Unit XL S, Temp module, CO<sub>2</sub> module, Heating Insert PS and CO<sub>2</sub> cover).

**Fish keratocytes.** Keratocytes were isolated from scales of freshly killed brook trout (*Salvelinus fontinalis*) as previously described<sup>35</sup> and imaged by phase contrast on an inverted Zeiss Axioscope using ×63 optics, and a halogen lamp as light source. Microinjection was performed with a micromanipulator (Leitz) and a micro-injector Femtojet (Eppendorf) controlling backpressure and injection pulses. Contours were analysed using the CellTrack software (Ohio State University).

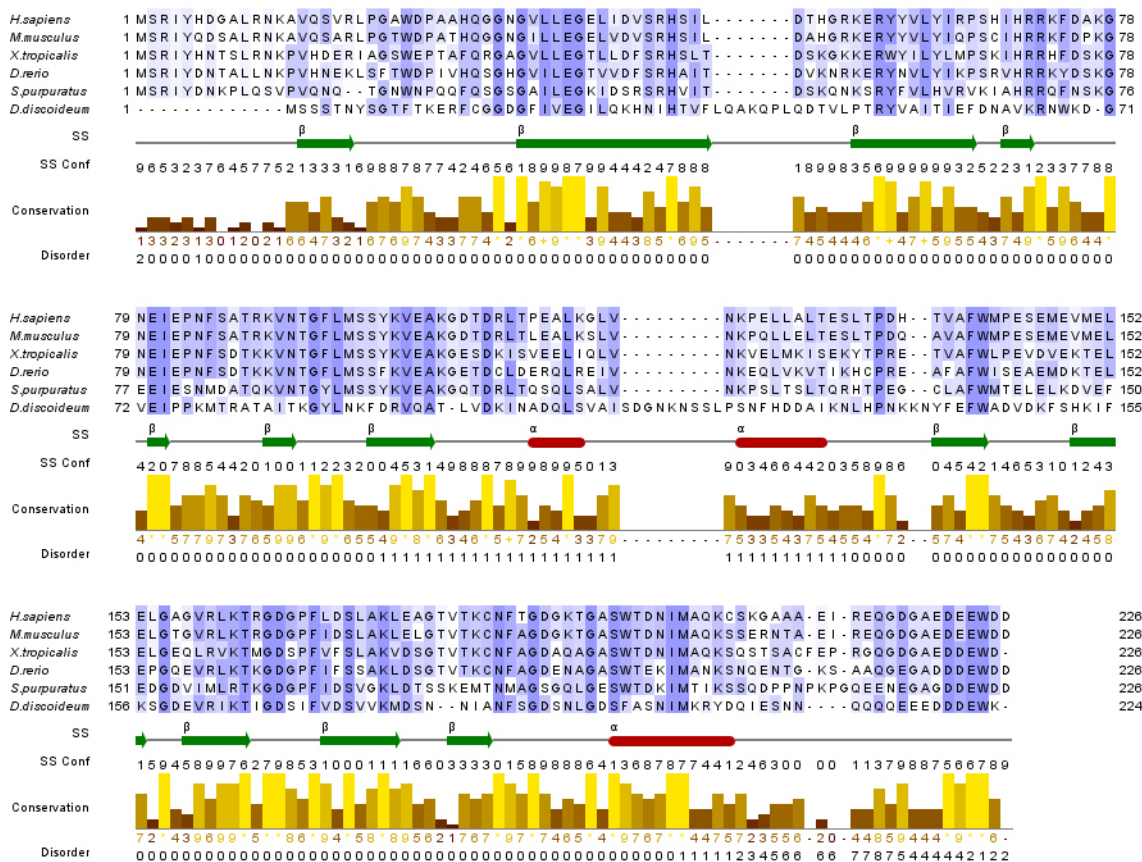
**Zebrafish.** Embryos were obtained by natural spawning of *Tg(-1.8gsc:GFP)ml1* fish<sup>36</sup>. In these embryos, prechordal plate cells can be identified by their expression of GFP. *In situ* hybridization was performed following standard protocols<sup>37</sup>. For loss of function experiments, a morpholino directed against *arpin* (GTTGTTCAT AAATACGACTCATCTTC, where the underlined anticodon corresponds to the initiating ATG codon), or a standard control morpholino (CCTCTTACCTCAG TTACAATTATA) was injected at the one-cell stage, together with histone2B-mCherry mRNAs or Lifeact-mCherry mRNAs, and GFP-Arpin mRNAs for rescue experiments. To analyse cell trajectories, confocal z-stacks were acquired every minute using a Nikon confocal spinning disk with an Evolve camera (Photometrics). Nuclei were tracked using Imaris (Bitplane). Further analyses were performed using custom routines in Matlab (MathWorks)<sup>38</sup>. All animal studies were done in accordance with the guidelines issued by the French Ministry of Agriculture (Decree no 2013-118) and have been submitted to Paris ethical committee no. 3. ***Dictyostelium discoideum.*** Cultivation and transformation by electroporation of *D. discoideum* cells was performed as described<sup>39</sup>. To knockout Arpin, a piece of genomic DNA containing the *arpin* coding sequence with its intron was amplified from Ax2 wild-type amoeba, using oligonucleotides DdArpin\_BU 5'-CGCGGATCCGCATGAGTTCAAGTACAAATTATAGT-3' and DdArpin\_SD 5'-CGCGTCGACTTTATTTCCATTCATCATCATCTTC-3'. The cloned PCR fragment was then used as a template to amplify a 5' fragment (using 5'-CGCGGATCCGCATGAGTTCAAGTACAAATTATAGT-3' and 5'-GCGCTGCAGCATCTGA AATTGCAACTGATAGTTG-3') and a 3' fragment (using 5'-GCGAAGCTTTC TTCTTTACCTTCAAATTTTCAT-3' and 5'-CGCGTCGACGTTGGTTATTT GATTCTATTTGATC-3'). These two fragments were cloned as to flank a cassette



carrying Blasticidin resistance in pLPBLP vector<sup>40</sup>. Arpin knockout clones were selected in HL5c-medium supplemented with  $10 \mu\text{g ml}^{-1}$  blasticidin S (Invivogen) after electroporation of the resulting vector. Recombination was assessed using diagnostic PCRs that distinguish knockout from wild-type amoeba. GFP-Arpin re-expressing knockout lines were obtained after electroporation of pDGFP-Arpin and selection with  $10 \mu\text{g ml}^{-1}$  geneticin (Sigma). Two time series with more than 30 cells each were acquired per amoeba. Two clones isolated after each transformation gave similar results.

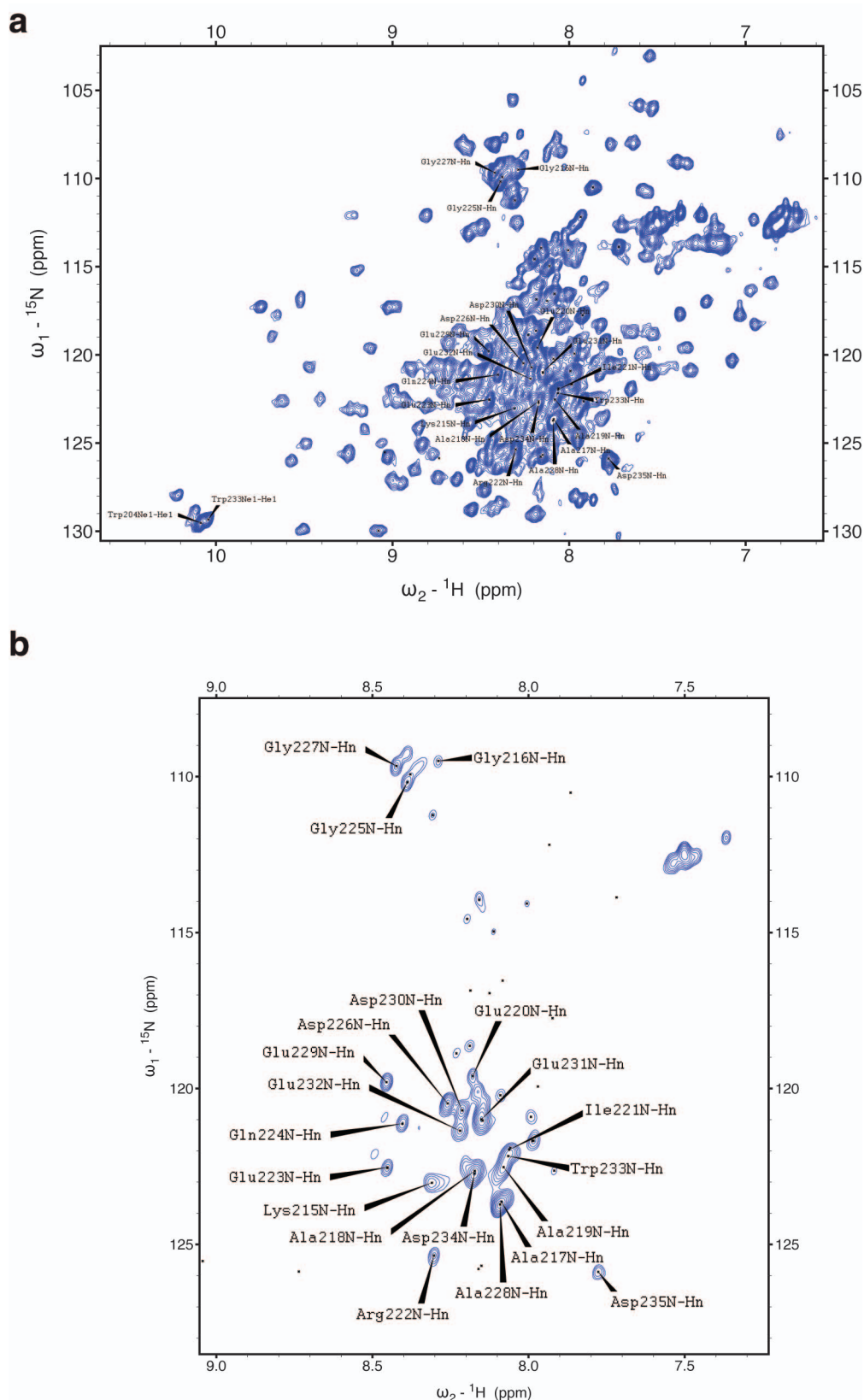
**Statistics.** Statistical analysis of the results was carried out with SigmaStat software (SPSS inc., v2.03). When data satisfied the two criteria of normality and equal variance, parametric tests were used: *t*-test to compare two groups; ANOVA for more than two. Where indicated, a bijective transformation was applied to the data to pass the two criteria of normality and equal variance. When data did not satisfy both criteria even after transformation, non-parametric tests were applied: Mann-Whitney to compare two groups; Kruskal-Wallis for more than two. A representative experiment is plotted and results are expressed as mean s.e.m. with respect to the number of cells (*n*). Differences were considered significant at confidence levels greater than 95% (two-tailed). Three levels of statistical significance are distinguished: \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001.

28. Lommel, S. *et al.* Actin pedestal formation by enteropathogenic *Escherichia coli* and intracellular motility of *Shigella flexneri* are abolished in N-WASP-defective cells. *EMBO Rep.* **2**, 850–857 (2001).
29. Millard, T. H., Behrendt, B., Launay, S., Futterer, K. & Machesky, L. M. Identification and characterisation of a novel human isoform of Arp2/3 complex subunit p16-ARC/ARPC5. *Cell Motil. Cytoskeleton* **54**, 81–90 (2003).
30. Dumontier, M., Hocht, P., Mintert, U. & Faix, J. Rac1 GTPases control filopodia formation, cell motility, endocytosis, cytokinesis and development in *Dictyostelium*. *J. Cell Sci.* **113**, 2253–2265 (2000).
31. Derivery, E. *et al.* Free Brick1 is a trimeric precursor in the assembly of a functional wave complex. *PLoS ONE* **3**, e2462 (2008).
32. Michelot, A. *et al.* Actin-filament stochastic dynamics mediated by ADF/cofilin. *Curr. Biol.* **17**, 825–833 (2007).
33. Machesky, L. M. *et al.* Scar, a WASp-related protein, activates nucleation of actin filaments by the Arp2/3 complex. *Proc. Natl Acad. Sci. USA* **96**, 3739–3744 (1999).
34. Marchand, J. B., Kaiser, D. A., Pollard, T. D. & Higgs, H. N. Interaction of WASP/Scar proteins with actin and vertebrate Arp2/3 complex. *Nature Cell Biol.* **3**, 76–82 (2001).
35. Urban, E., Jacob, S., Nemethova, M., Resch, G. P. & Small, J. V. Electron tomography reveals unbranched networks of actin filaments in lamellipodia. *Nature Cell Biol.* **12**, 429–435 (2010).
36. Doitsidou, M. *et al.* Guidance of primordial germ cell migration by the chemokine SDF-1. *Cell* **111**, 647–659 (2002).
37. Hauptmann, G. & Gerster, T. Two-color whole-mount *in situ* hybridization to vertebrate and *Drosophila* embryos. *Trends Genet.* **10**, 266 (1994).
38. Dumortier, J. G., Martin, S., Meyer, D., Rosa, F. M. & David, N. B. Collective mesendoderm migration relies on an intrinsic directionality signal transmitted through cell contacts. *Proc. Natl Acad. Sci. USA* **109**, 16945–16950 (2012).
39. Schirenbeck, A., Bretschneider, T., Arasada, R., Schleicher, M. & Faix, J. The Diaphanous-related formin dDia2 is required for the formation and maintenance of filopodia. *Nature Cell Biol.* **7**, 619–625 (2005).
40. Faix, J., Kreppel, L., Shaulsky, G., Schleicher, M. & Kimmel, A. R. A rapid and efficient method to generate multiple gene disruptions in *Dictyostelium discoideum* using a single selectable marker and the Cre-loxP system. *Nucleic Acids Res.* **32**, e143 (2004).
41. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
42. Clamp, M., Cuff, J., Searle, S. M. & Barton, G. J. The Jalview Java alignment editor. *Bioinformatics* **20**, 426–427 (2004).
43. McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405 (2000).
44. Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F. & Jones, D. T. The DISOPRED server for the prediction of protein disorder. *Bioinformatics* **20**, 2138–2139 (2004).
45. Blanchoin, L. *et al.* Direct observation of dendritic actin filament networks nucleated by Arp2/3 complex and WASP/Scar proteins. *Nature* **404**, 1007–1011 (2000).
46. Montero, J. A., Kilian, B., Chan, J., Bayliss, P. E. & Heisenberg, C. P. Phosphoinositide 3-kinase is required for process outgrowth and cell polarization of gastrulating mesendodermal cells. *Curr. Biol.* **13**, 1279–1289 (2003).



**Extended Data Figure 1 | Prediction of secondary structure elements and disordered regions of Arpin.** A multiple alignment of the Arpin orthologues was performed with MUSCLE<sup>41</sup> and displayed with Jalview<sup>42</sup>. Two methods relying on multiple alignments of Arpin orthologues were used to predict secondary structures and disordered regions, Psipred<sup>43</sup> and Disopred<sup>44</sup>, respectively. The predicted secondary structure (SS) elements are indicated by green arrows for  $\beta$ -strands, red cylinders for  $\alpha$ -helices, and a black line for coils;

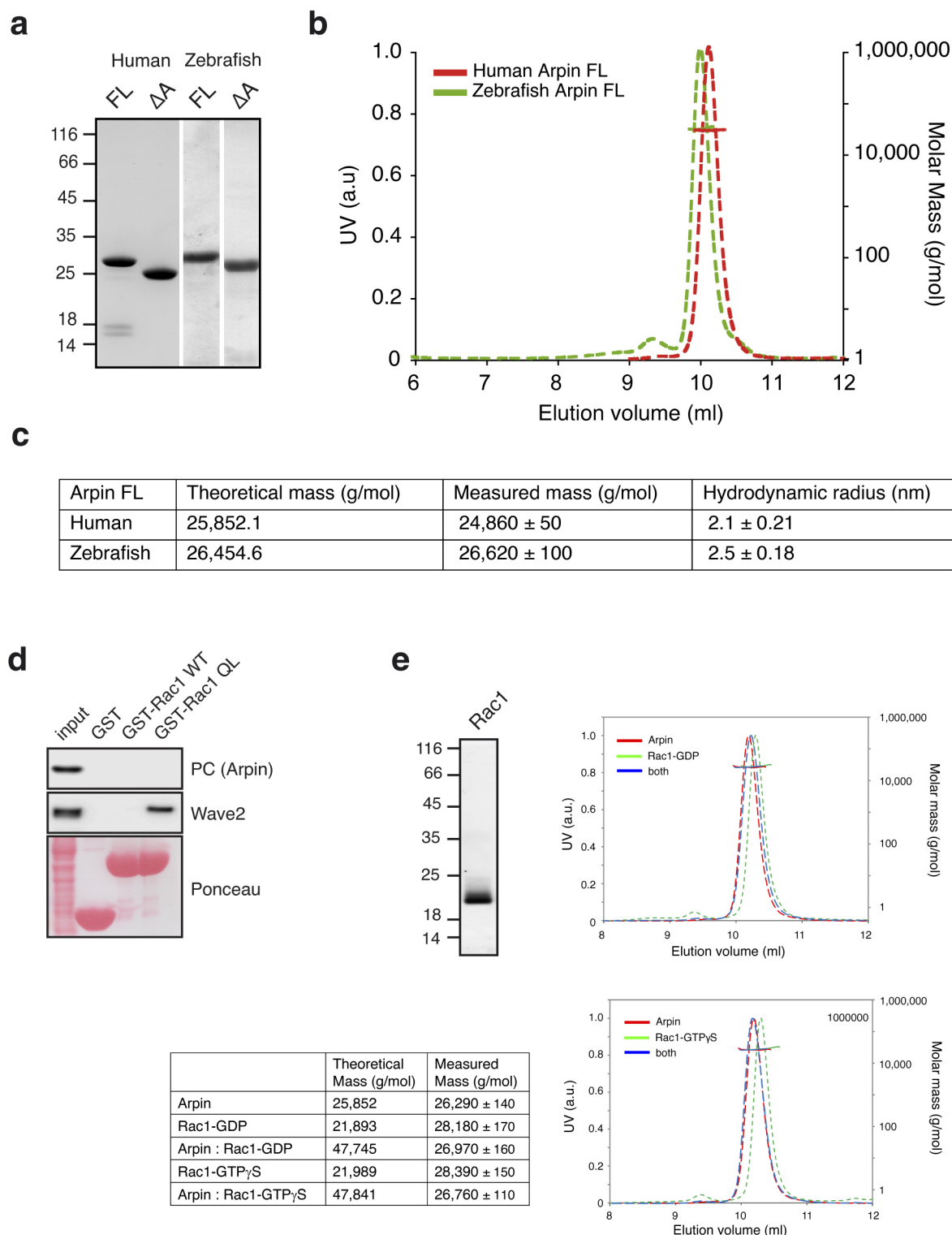
the associated confidence (conf) score is displayed below, ranging from 0 to 9 for poor and high confidence, respectively. Amino acid conservation is indicated by a 0 to 10 score, and highlighted by brown to yellow histogram bars. The confidence in predicting disorder is also scored from 0 to 10, by multiplying tenfold the Disopred probability. Arpin is predicted to be a structured protein with the notable exception of the 20 C-terminal residues, which are predicted to be disordered with high confidence.



**Extended Data Figure 2** | NMR analysis of  ${}^{15}\text{N}$ -labelled human Arpin. Both views represent  ${}^1\text{H}$ - ${}^{15}\text{N}$  heteronuclear single quantum coherence (HSQC) spectra. Each peak corresponds to the  ${}^1\text{H}$ - ${}^{15}\text{N}$  backbone amide bond of a specific residue. The position of a  ${}^1\text{H}$ - ${}^{15}\text{N}$  peak in the spectrum depends on the chemical environment of the corresponding residue. **a**, Such a scattered distribution of peaks is characteristic of a folded protein. The last 20 residues were assigned to individual peaks and are displayed on the spectrum. These

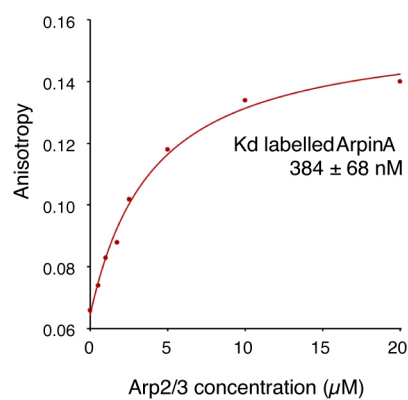
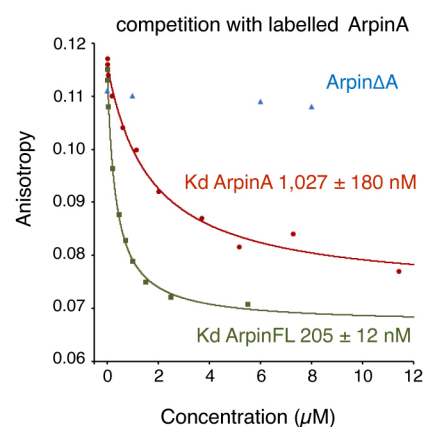
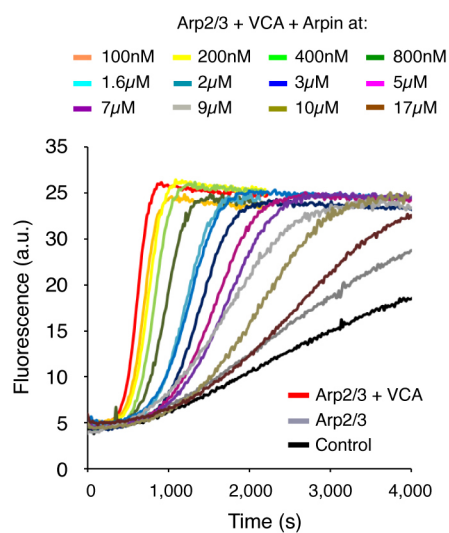
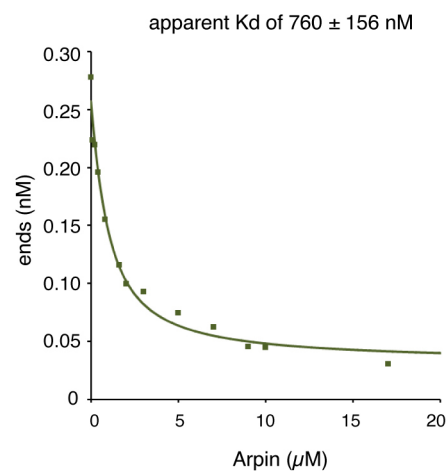
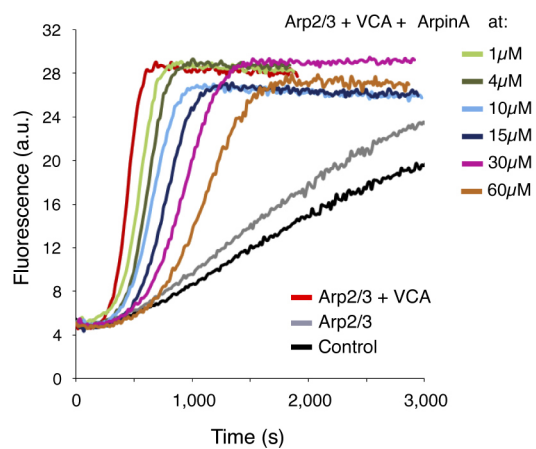
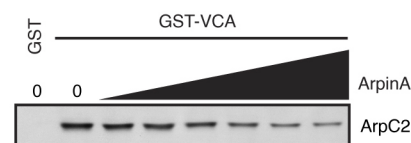
residues are clustered in the centre of the spectrum. **b**, Same HSQC spectrum displayed with a higher threshold to display only high peaks. The height of a peak depends on the mobility of the residue on a picosecond to millisecond timescale. This spectrum experimentally demonstrates that the 20 C-terminal residues are highly mobile. This result confirms that, as predicted, the Arp2/3-binding site of Arpin is exposed as a poorly structured tail of the protein.





**Extended Data Figure 3 | Characterization of recombinant Arpin.** **a**, Full-length Arpin or Arpin $\Delta$ A from human or zebrafish cDNA was expressed in *E. coli* and purified. Purity was assessed by SDS-PAGE and coomassie staining. These proteins were used for *in vitro* actin polymerization assays and for fish keratocyte injection, respectively. **b**, Analysis of the molar mass of full-length Arpins by size-exclusion chromatography coupled to multiangle light scattering (SEC-MALS). The ultraviolet measurement (left axis, dashed line) and the molar mass (right axis, horizontal solid line) were plotted as a function of column elution volume. **c**, SEC-MALS measures of masses indicate that both proteins are monomeric in solution. **d**, GST pull-down using a lysate of 293 cells overexpressing PC-tagged Arpin and purified GST-Rac1 wild type,

GST-Rac1(Gln61Leu) or GST alone as a negative control. Arpin did not associate with either type of Rac. By contrast, the endogenous WAVE complex bound to Rac(Gln61Leu), but not Rac wild type, as expected from a Rac effector. **e**, Untagged Rac1 was purified from *E. coli* and then loaded with either GDP or GTP $\gamma$ S. Human Arpin (60  $\mu$ M), Rac1 (120  $\mu$ M) and mixture of these two proteins were analysed by SEC-MALS as above. SEC was run in 20 mM HEPES, 100 mM NaCl, pH 7.4. The height of ultraviolet peaks was normalized to 1 to be displayed on the same figure. A single peak was detected in all cases. The measured masses indicate that no complex is formed between Arpin and Rac and that the single peak observed in the mixture corresponds to cofractionation of the two proteins of similar mass by SEC.

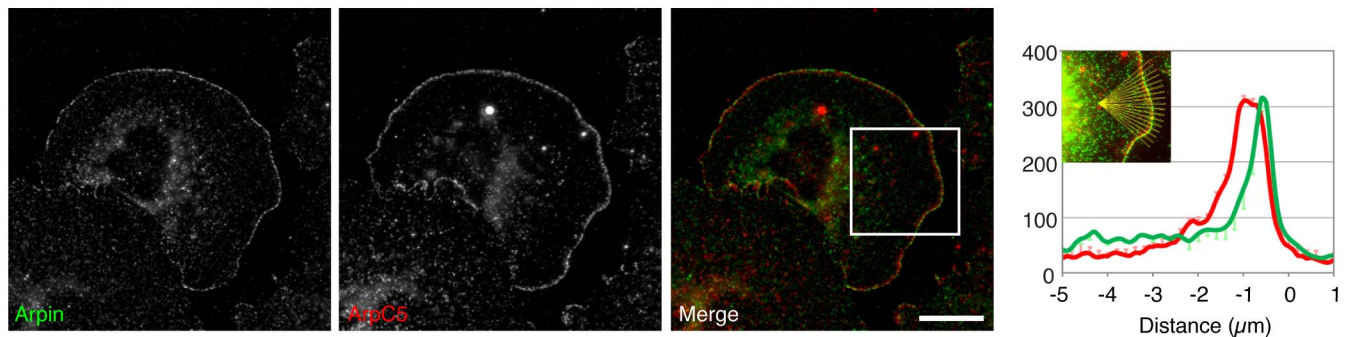
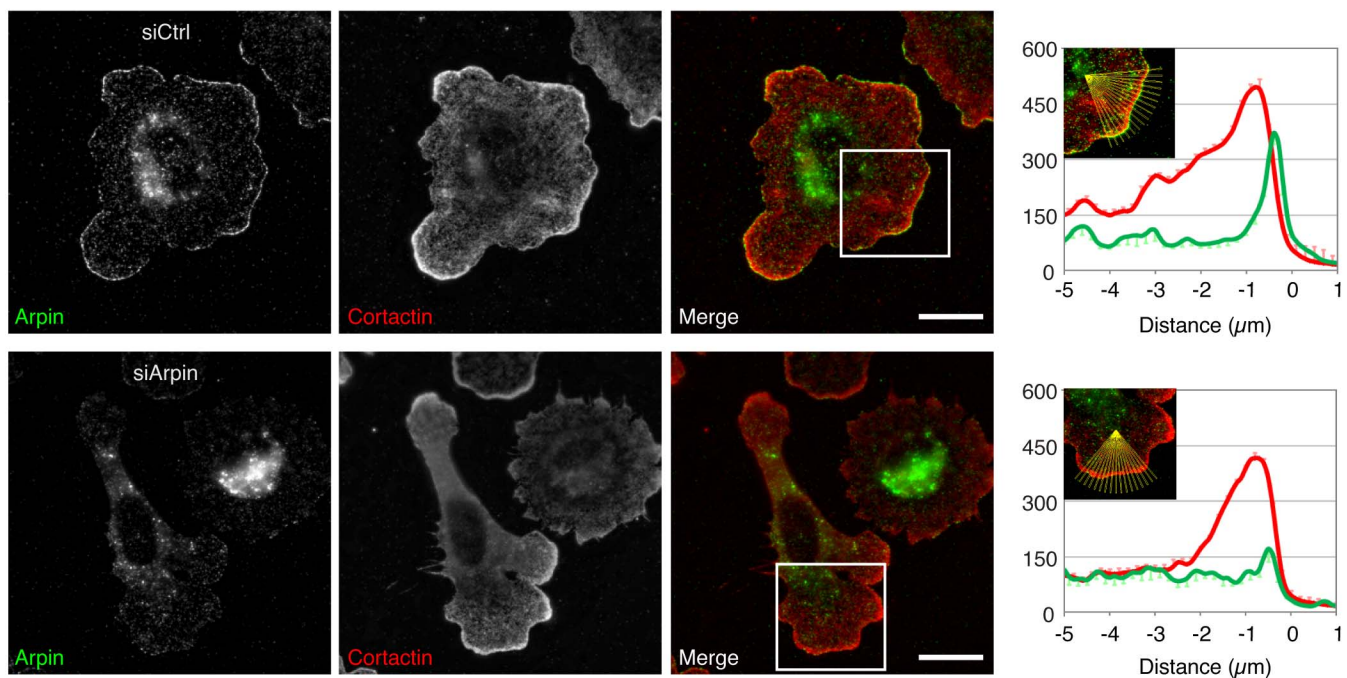
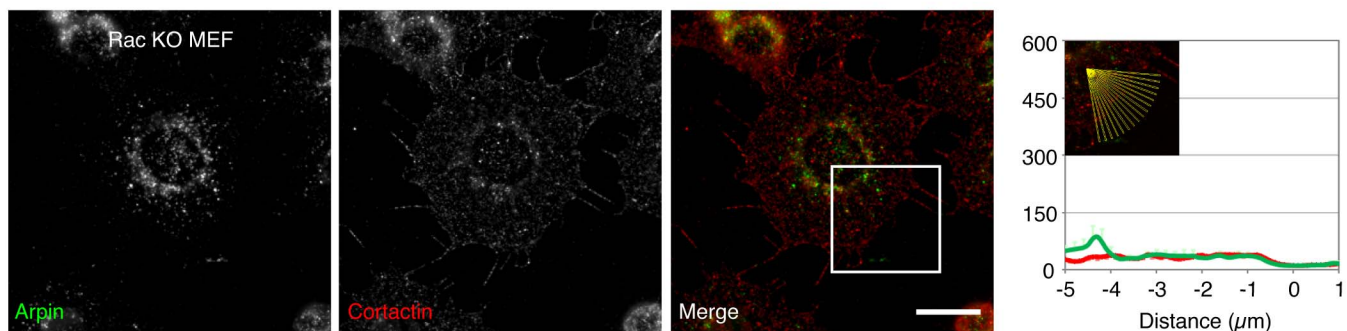
**a****b****c****d****e****f**

**Extended Data Figure 4 | Arpin directly binds to the Arp2/3 complex.**

**a**, Fluorescence anisotropy measurements of labelled ArpinA peptide binding at equilibrium to the purified Arp2/3 complex at the indicated concentrations. **b**, Labelled ArpinA peptide bound to the Arp2/3 complex was then titrated with purified Arpin full-length, Arpin $\Delta$ A or unlabelled ArpinA peptide as indicated. Full-length Arpin displaces the labelled ArpinA peptide more efficiently than the A peptide. Arpin $\Delta$ A is unable to displace the ArpinA peptide. Curves that best fit the values yield the indicated equilibrium constants. **c**, Arpin inhibits Arp2/3 activation in the pyrene-actin assay. Part of this experiment is displayed in Fig. 1c, more curves are plotted here. **d**, From curves in c, the number of actin

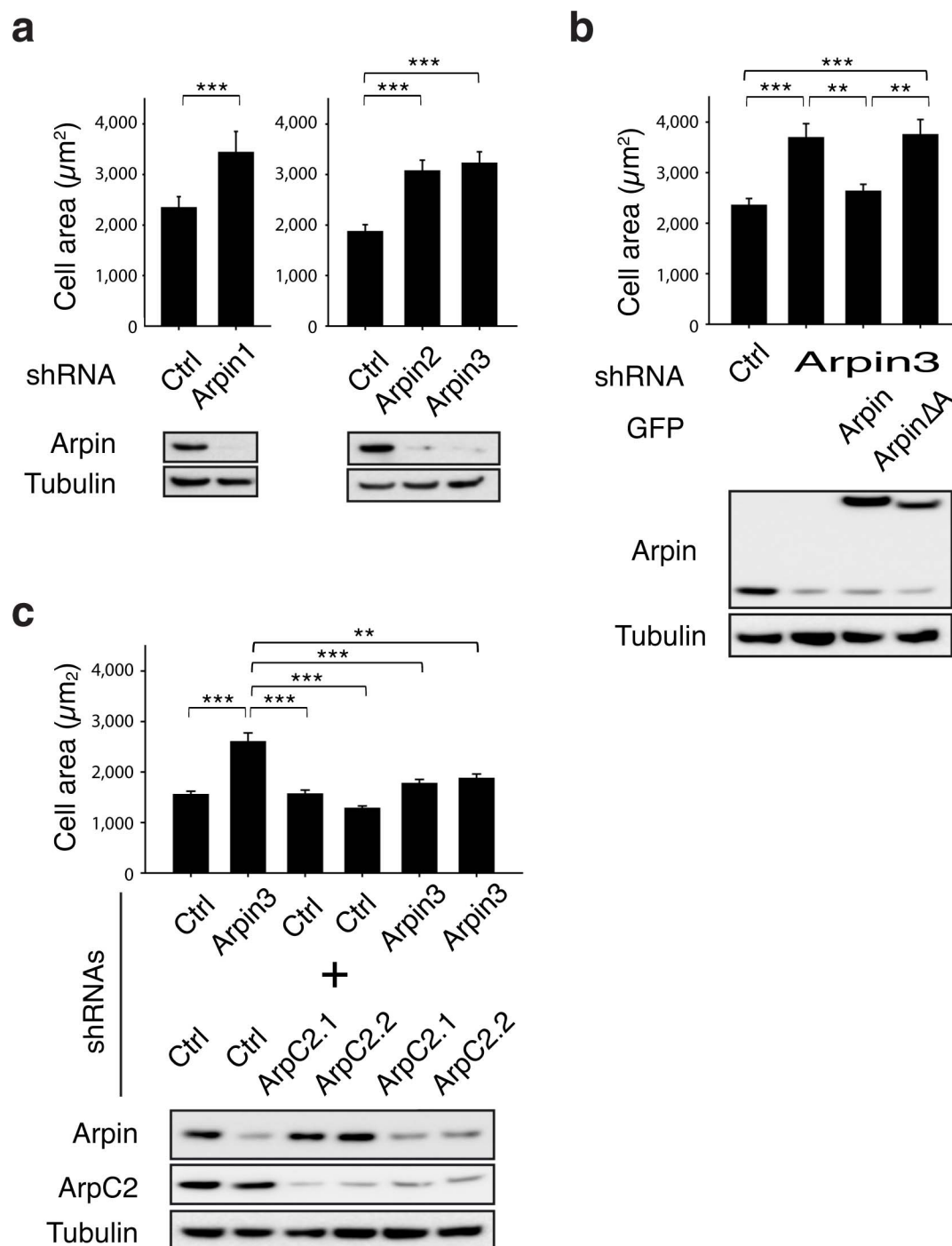
barbed ends is calculated from the slope at half-polymerization using the relationship described previously<sup>45</sup>. Best fit of the values indicate an apparent  $K_d$  value of  $760 \pm 156$  nM for the Arp2/3 complex in a mixture including actin and the VCA. **e**, ArpinA inhibits Arp2/3 activation in a dose dependent manner in the pyrene-actin assay. Conditions: 2  $\mu$ M actin (10% pyrene-labelled), 500 nM VCA, 20 nM Arp2/3 and ArpinA at the indicated concentrations. **f**, ArpinA competes with the NPF for Arp2/3 binding. Arp2/3 is displaced from its interaction with 5  $\mu$ M GST N-WASP VCA immobilized on glutathione beads by the Arpin acidic peptide (304  $\mu$ M and serial twofold dilutions).



**a****b****c**

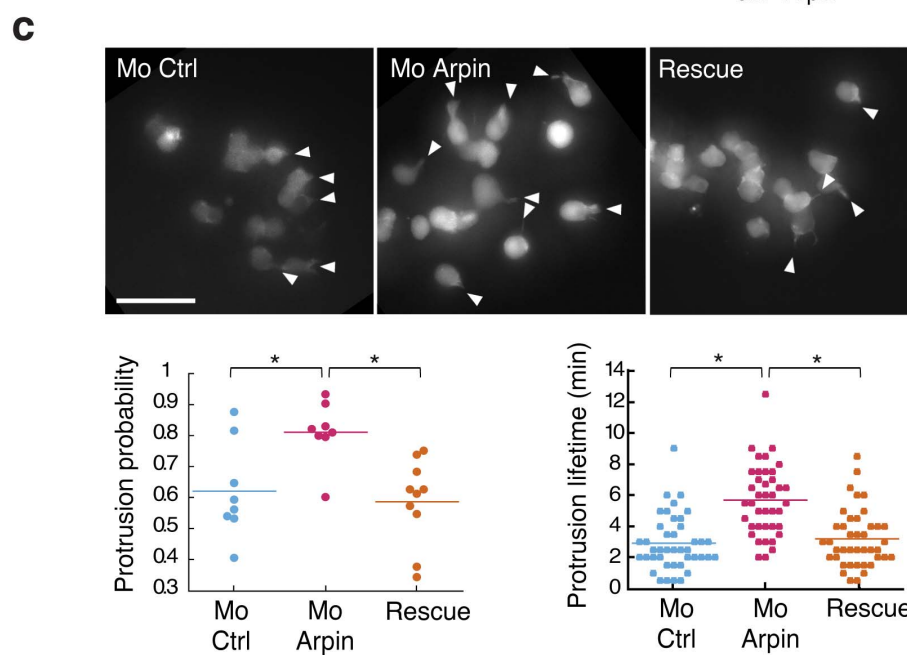
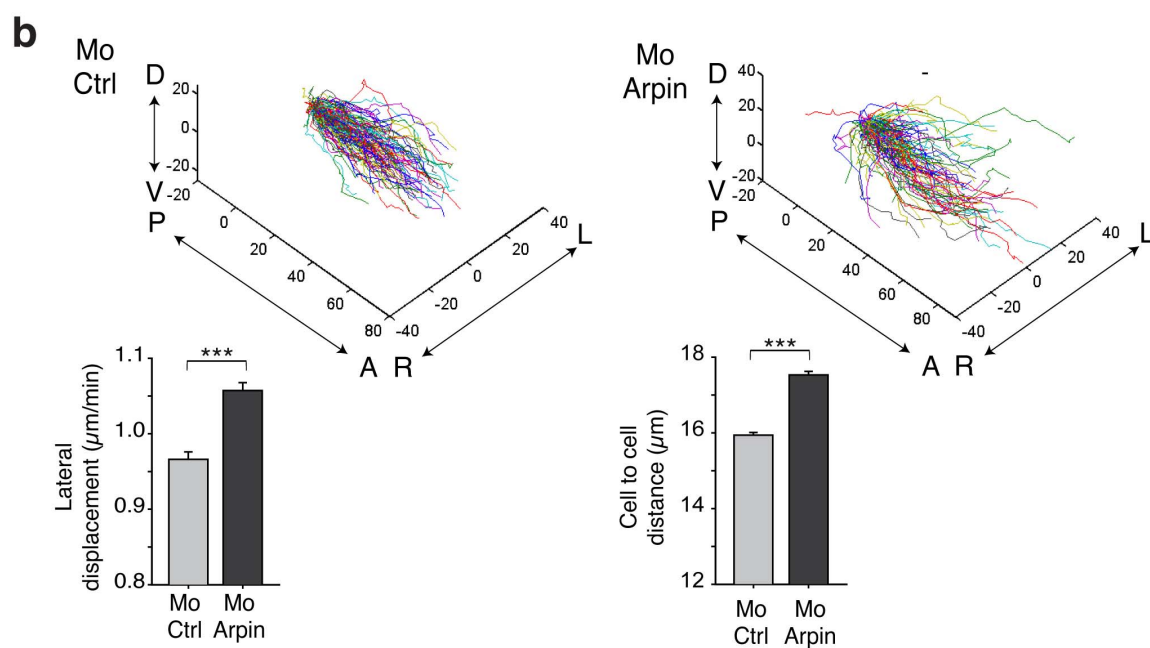
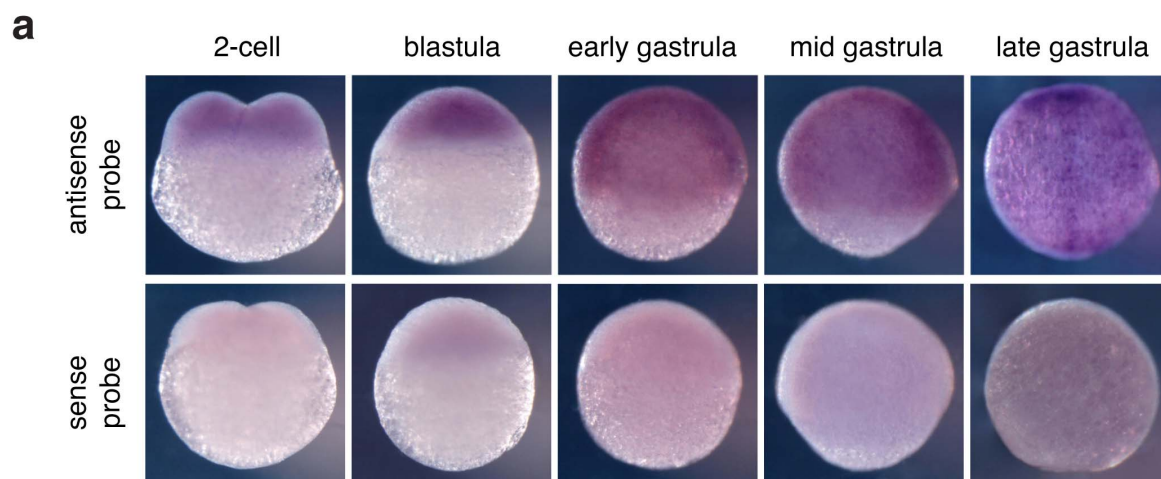
**Extended Data Figure 5 | Specific localization of Arpin at the lamellipodium tip.** **a**, Arpin overlaps with the Arp2/3 complex at the lamellipodium tip. **b**, Arpin overlaps with cortactin at the lamellipodium tip. Arpin staining is lost after short interfering RNA (siRNA)-mediated depletion. Intensity profiles along multiple line scans encompassing the cell periphery were registered to the outer edge of the staining of a lamellipodial marker. This marker was Arpin in **a** and cortactin in **b**. The multiple line scans were then averaged and displayed as an intensity plot, where the  $y$  axis represents fluorescent intensity, arbitrary units (mean  $\pm$  s.e.m.,  $n = 17$ , 16 and 17, respectively). Scale bar, 20  $\mu$ m. Arp2/3 localization extends rearwards relative to Arpin localization. This result is because the Arp2/3 complex becomes a branched junction when activated by the WAVE complex at the lamellipodium

tip. The branched junction undergoes retrograde flow like actin itself due to actin filament elongation<sup>9,12</sup>. Cortactin recognizes Arp2/3 at the branch junction and is thought to stabilize branched actin networks<sup>13</sup>. As a marker of the branched junction, cortactin stains the width of lamellipodia, like the Arp2/3 complex. **c**, Rac1 knockout MEFs that lack lamellipodia<sup>14</sup> are completely devoid of Arpin staining at the cell periphery, in line with the complete absence of lamellipodia indicated here by the absence of cortactin staining. Arpin is normally expressed in the Rac1 knockout MEFs (see Fig. 2c). Intensity profiles along multiple line scans encompassing the cell periphery were averaged after manual drawing of the cell edge (mean  $\pm$  s.e.m.,  $n = 16$ ). Scale bar, 20  $\mu$ m.



**Extended Data Figure 6 | Arpin regulates cell spreading through its interaction with the Arp2/3 complex.** Arpin was depleted from human RPE1 cells after transient transfection of shRNA plasmids and blasticidin-mediated selection of transfected cells. After 5 days, cells were either analysed by western blot or used for the spreading assay. Cells were serum-starved for 90 min in suspension in polyHEMA-coated dishes and then allowed to spread on collagen-I-coated coverslips for 2 h. Phalloidin staining was used to calculate cell surface area of individual cells using ImageJ. Mean  $\pm$  s.e.m.; \*\* $P$  < 0.01, \*\*\* $P$  < 0.001;  $t$ -test or ANOVA when more than two conditions. **a**, Arpin

depletion increases cell spreading ( $n$  = 57 and 52, respectively). The same effect is obtained with three shRNAs targeting Arpin ( $n$  = 51, 48 and 59, respectively). **b**, This effect is rescued by GFP–Arpin expression in knockdown cells, but not by GFP–ArpinΔA expression ( $n$  = 63, 56, 63 and 52, respectively). **c**, Combined depletion of Arpin and the Arp2/3 complex reverses the phenotype of Arpin depletion. The effect is seen with two shRNAs targeting ArpC2 ( $n$  = 56, 60, 69, 68, 63 and 66, respectively). The last two experiments indicate that Arpin exerts its effect on cell spreading through its ability to regulate the Arp2/3 complex.

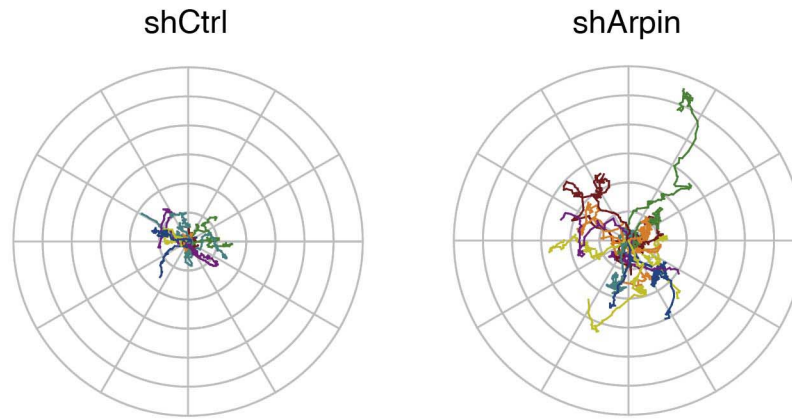
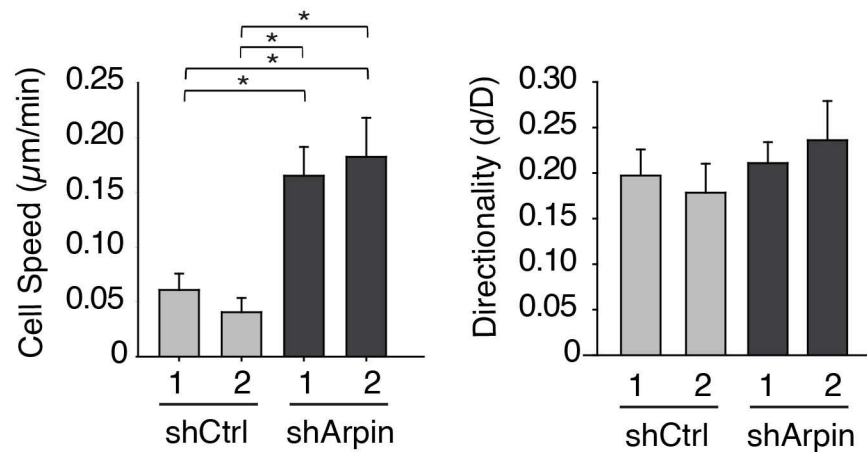




**Extended Data Figure 7 | Arpin regulates protrusion frequency of prechordal plate cells and their collective migration in zebrafish embryos.**

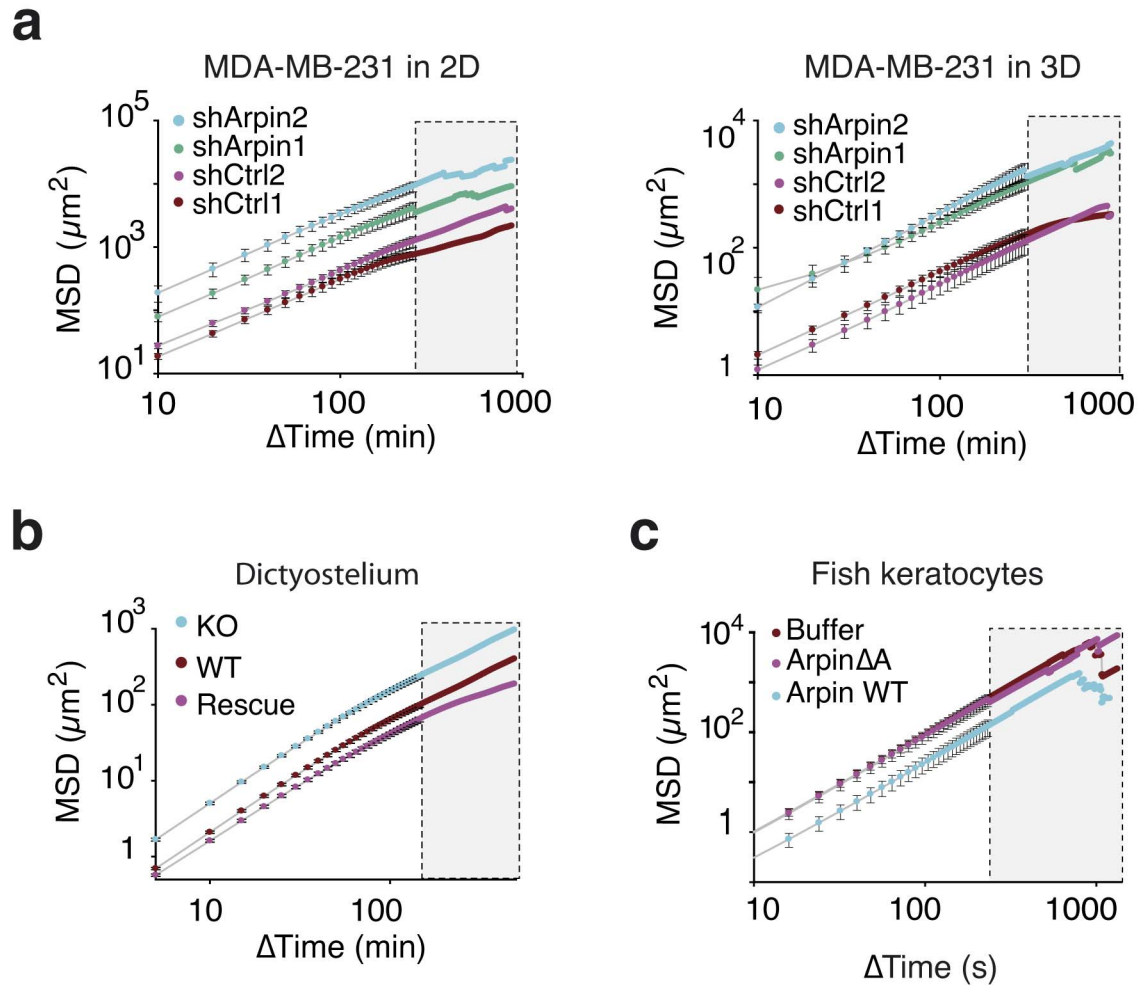
**a**, *In situ* hybridization of *arpin* probe in zebrafish embryos at different stages. *arpin* mRNAs are maternally deposited. During gastrulation, *arpin* is expressed in hypoblast, which includes the prechordal plate. **b**, Three-dimensional trajectories of prechordal plate cells in embryos injected with control or *arpin* morpholino. During fish gastrulation, prechordal plate cells migrate collectively in a straight direction from the margin of the embryo towards the animal pole<sup>38,46</sup>. Loss of *arpin* function induces dispersion as evidenced by increased lateral cell displacement ( $n = 1,516$  and  $1,546$ ) and a higher distance between cells ( $n = 194$  and  $235$ ). Lateral displacement is the cell movement perpendicular to main direction of the migration. Distance between cells refers to the average distance of the nucleus of a given cell to the nuclei of its five closest neighbours. Mean  $\pm$  s.e.m.; \*\*\* $P < 0.001$ ,  $t$ -test. **c**, At the onset of

gastrulation prechordal plate cells derived from morpholino injected embryos were transplanted into the prechordal plate of an untreated host embryo at the same stage in order to allow imaging of cell autonomous effects on protrusion formation. Donor embryos are injected with control or *arpin* morpholinos and mRNAs encoding Lifeact-mCherry as well as GFP-Arpin for the rescue. Time-lapse imaging of injected cells is performed by epifluorescence to reveal Lifeact, a marker of filamentous actin, which stains actin-based protrusions. For each cell, presence of a protrusion was assessed at each frame to deduce probability of protrusion presence and protrusion lifetimes. *arpin* loss of function increases the probability of presence of protrusions ( $n = 8, 8$  and  $10$ , respectively; \* $P < 0.05$ , ANOVA) and their duration (in this case,  $n$  corresponds to the number of protrusions ( $n = 42, 41$  and  $40$ , respectively; \* $P < 0.05$ , Kruskal–Wallis). Protrusions are indicated by arrowheads. Scale bar,  $50\ \mu\text{m}$ .

**a****b**

**Extended Data Figure 8 | Arpin depletion increases cell migration in three dimensions.** Stable MDA-MB-231 clones depleted of Arpin or not were embedded in a collagen gel. **a**, Single-cell trajectories illustrate that control cells hardly move in this dense environment (see Supplementary Video 4), unlike Arpin-depleted cells, which explore a significant territory, albeit at lower pace than in two dimensions, as evidenced by mean square displacement (Extended

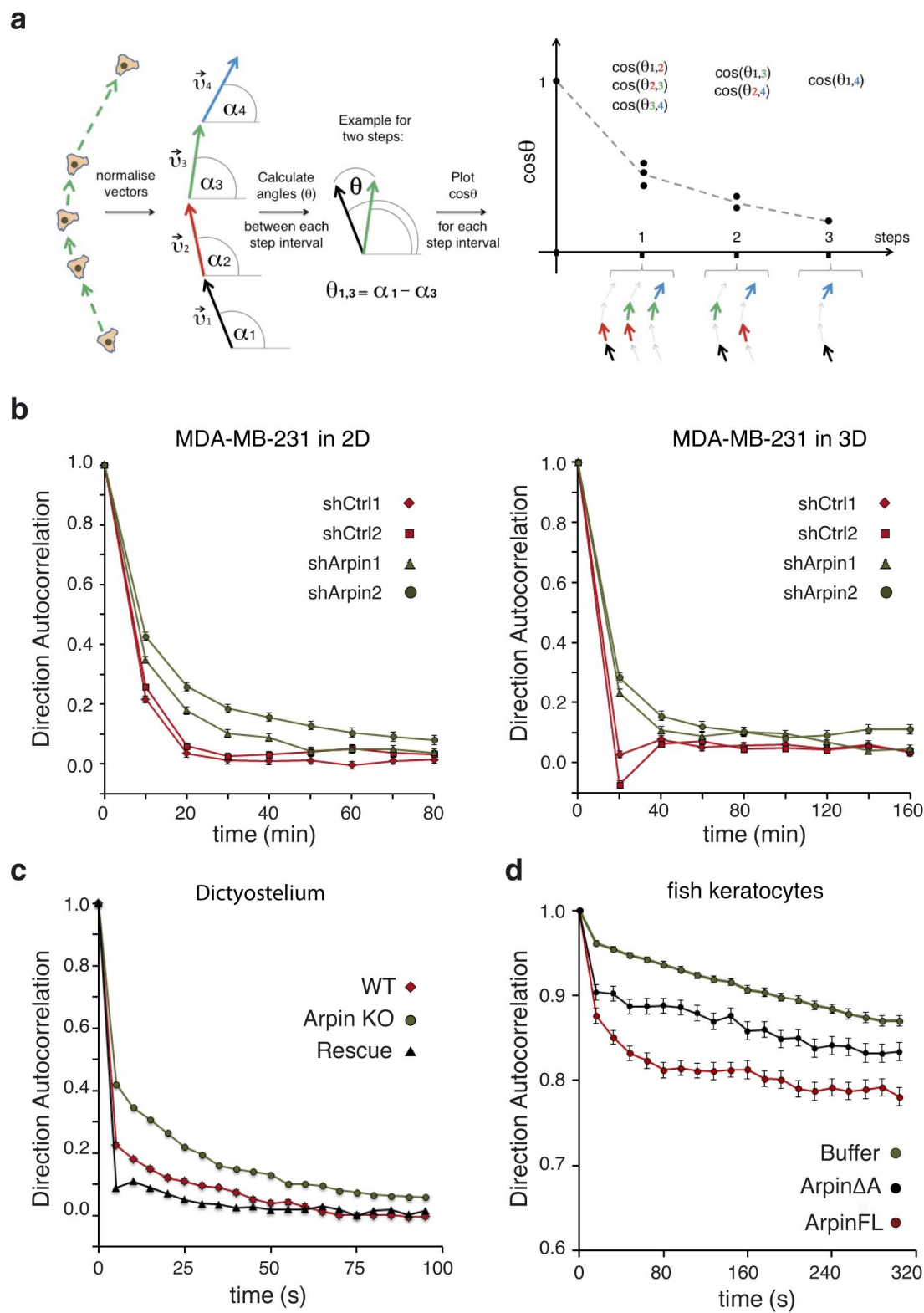
Data Fig. 9). **b**, Cell speed is significantly increased in the Arpin-depleted clones. Mean  $\pm$  s.e.m.;  $n = 27, 25, 26$  and  $17$ , respectively,  $*P < 0.05$ , Kruskal–Wallis, two experiments. Directional persistence, calculated by  $d/D$ , is not significantly different in the clones depleted of Arpin or not. Direction autocorrelation (Extended Data Fig. 10), however, shows an increased directionality in the Arpin-depleted cells at the earliest time points.



**Extended Data Figure 9 | Analysis of mean square displacement of the different migration experiments.** The mean square displacement gives a measure of the area explored by cells for any given time interval. By setting a positional vector on the cellular trajectory at time  $t$ , the MSD is defined as:  $\text{MSD}(\Delta t) = \langle [x(t+t_0) - x(t_0)]^2 + [y(t+t_0) - y(t_0)]^2 \rangle_{t,N}$ , in which brackets  $\langle \rangle$  indicate averages over all starting times  $t_0$  and all cells  $N$ . For each time interval  $\Delta\text{time}$ , mean and s.e.m. are plotted. Error bars corresponding to s.e.m. are plotted, even if too small to be visible. The grey area excludes the noisy part of curves corresponding to large time intervals where less data points are available. **a**, MDA-MB-231 depleted or not of Arpin in a two- or three-dimensional

environment. Arpin-depleted MDA-MB-231 cells explore a larger territory than the controls in time intervals examined (for two dimensions,  $n$  as indicated in Fig. 3a;  $P < 0.001$ , two-way ANOVA with time and conditions; for three dimensions,  $n$  as indicated in Extended Data Fig. 8;  $P < 0.001$ , two-way ANOVA with time and conditions). **b**, *Dictyostelium discoideum* knockout amoebae explore a larger territory than controls and rescued amoebae ( $n$  as indicated in Fig. 3b,  $P < 0.001$ , two-way ANOVA with time and conditions). **c**, Arpin-injected fish keratocytes explore a smaller territory than the controls ( $n$  as indicated in Fig. 4b;  $P < 0.001$ , two-way ANOVA with time and conditions).





**Extended Data Figure 10 | Analysis of direction autocorrelation of the different migration experiments.** **a**, Principle of the analysis. A hypothetical cell trajectory is depicted. Each step is represented by a vector of normalized length.  $\theta$  is the angle between compared vectors. The plot illustrates the  $\cos\theta$  values for the putative trajectory of four steps (colour-coded). Averaging these  $\cos\theta$  values yields the direction autocorrelation (DA) function of time that measures the extent to which these vectors are aligned over different time intervals. The DA function is defined as:  $DA(t) = \langle v(t_0) \cdot v(t_0 + t) \rangle_{t_0, N} = \langle \cos \theta(t_0, t_0 + t) \rangle_{t_0, N}$ , in which  $v(t_0)$  is the vector at the starting time  $t_0$ , and  $v(t_0 + t)$  the vector at  $t_0 + t$ . Brackets indicate that all calculated cosines are averaged for all possible starting times ( $t_0$ ) over all cells ( $N$ ). For each time interval  $t$ , vectors from all cell trajectories were

used to compute average and sem. Error bars corresponding to s.e.m. are plotted, even if too small to be visible. **b**, Arpin-depleted MDA-MB-231 clones turn less than control cells (for two dimensions,  $n$  as indicated in Fig. 3a;  $P < 0.05$  between 10 and 40 min, Kruskal–Wallis; for three dimensions,  $n$  as indicated in Extended Data Fig. 8;  $P < 0.05$  at time 10 min, Kruskal–Wallis). **c**, Arpin knockout amoebae turn less than wild-type amoebae, and GFP–Arpin overexpressing knockout amoebae (rescue) turn more than wild type ( $n$  as indicated in Fig. 3b;  $P < 0.05$  between 5 and 85 s, Kruskal–Wallis). **e**, Arpin-injected fish keratocytes turn more than buffer-injected cells, and Arpin $\Delta A$ -injected keratocytes turn more than buffer-injected but less than full-length-Arpin-injected keratocytes ( $n$  as indicated in Fig. 4b;  $P < 0.05$  between 16 and 272 s, Kruskal–Wallis).

# Directional tissue migration through a self-generated chemokine gradient

Erika Donà<sup>1</sup>, Joseph D. Barry<sup>1</sup>, Guillaume Valentin<sup>1†</sup>, Charlotte Quirin<sup>1</sup>, Anton Khmelinskii<sup>2</sup>, Andreas Kunze<sup>1</sup>, Sevi Durdu<sup>1</sup>, Lionel R. Newton<sup>1</sup>, Ana Fernandez-Minan<sup>1‡</sup>, Wolfgang Huber<sup>1</sup>, Michael Knop<sup>2</sup> & Darren Gilmour<sup>1</sup>

The directed migration of cell collectives is a driving force of embryogenesis<sup>1–3</sup>. The predominant view in the field is that cells in embryos navigate along pre-patterned chemoattractant gradients<sup>2</sup>. One hypothetical way to free migrating collectives from the requirement of long-range gradients would be through the self-generation of local gradients that travel with them<sup>4,5</sup>, a strategy that potentially allows self-determined directionality. However, a lack of tools for the visualization of endogenous guidance cues has prevented the demonstration of such self-generated gradients *in vivo*. Here we define the *in vivo* dynamics of one key guidance molecule, the chemokine Cxcl12a, by applying a fluorescent timer approach to measure ligand-triggered receptor turnover in living animals. Using the zebrafish lateral line primordium as a model, we show that migrating cell collectives can self-generate gradients of chemokine activity across their length via polarized receptor-mediated internalization. Finally, by engineering an external source of the atypical receptor Cxcr7 that moves with the primordium, we show that a self-generated gradient mechanism is sufficient to direct robust collective migration. This study thus provides, to our knowledge, the first *in vivo* proof for self-directed tissue migration through local shaping of an extracellular cue and provides a framework for investigating self-directed migration in many other contexts including cancer invasion<sup>6</sup>.

The zebrafish lateral line primordium is a migrating collective of approximately 100 cells that deposits a series of mechanosensory organs along the main body axis<sup>7</sup> and whose route is defined by a stripe of the chemokine Cxcl12a (also known as Sdf1a)<sup>8</sup>. Previous genetic studies have shown that the primordium can follow the Cxcl12a stripe efficiently in both directions<sup>9</sup>, excluding an instructive role for a pre-patterned chemokine gradient. Two ‘self-generated gradient’ hypotheses, whereby tissues sculpt directional gradients from uniform extrinsic guidance cues, have been proposed to explain this remarkable behaviour. First, Cxcl12a internalization by the guidance receptor Cxcr4b could generate a dynamic chemokine gradient at the tissue’s leading edge, which in mathematical models is sufficient to drive forward migration<sup>10</sup>. Second, directional symmetry could be broken by internalization of Cxcl12a at the tissue’s trailing edge by Cxcr7 (refs 4, 11), a receptor with known Cxcl12a sinking<sup>12,13</sup> and signalling<sup>14,15</sup> functions. However, a lack of tools for the visualization of endogenous guidance signal activity has prevented the validation of self-generated gradient models in any *in vivo* context.

We therefore established a novel readout of chemokine activity by measuring receptor turnover at the cell surface, a well-characterized cellular response to ligand binding<sup>16</sup>, using a tandem fluorescent protein timer (tFT) approach<sup>17</sup>. To this aim, we tagged the zebrafish Cxcr4b receptor with a tFT consisting of the superfolder green fluorescent protein (sfGFP) coupled to the slower maturing red fluorescent protein TagRFP. The red/green fluorescence ratio at the plasma membrane, hereafter referred to as the ‘lifetime ratio’, provides a direct indicator of the age of protein populations, whereby newly synthesized proteins

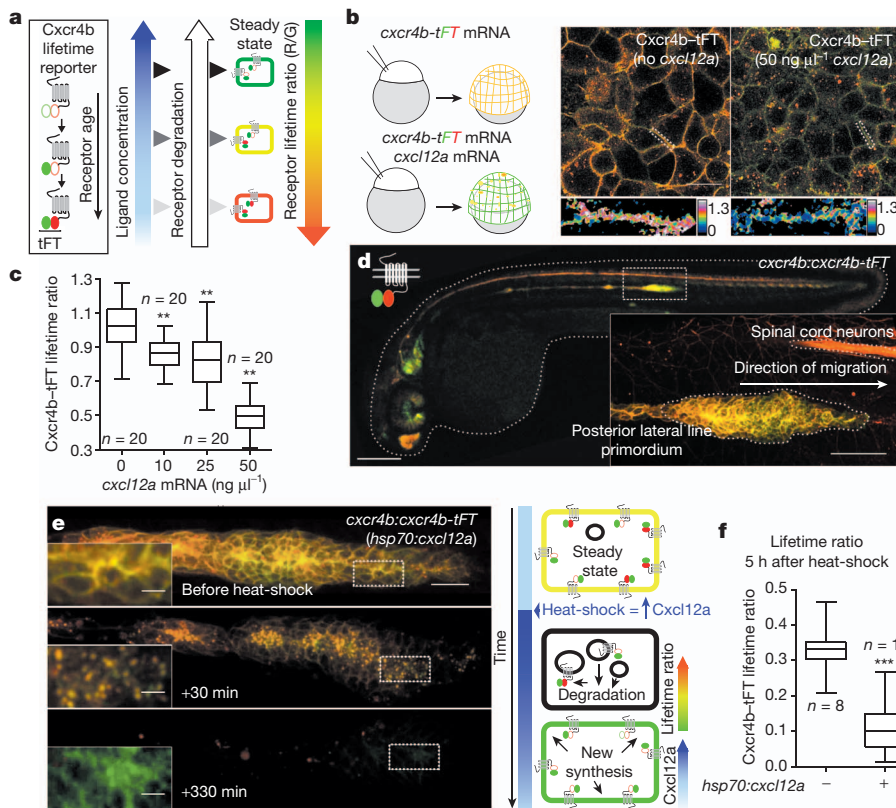
show only green fluorescence, with red fluorescence increasing progressively with protein age. As the rate of ligand-triggered internalization determines how long receptors reside at the cell surface, receptor lifetime ratios would be predicted to decrease as chemokine levels increase (Fig. 1a). We first tested the response of the lifetime reporter in early-stage zebrafish embryos injected with defined concentrations of *cxcl12a* mRNA. This revealed that Cxcr4b–tFT lifetime ratios at the plasma membrane progressively decreased over a range of *cxcl12a* levels (Fig. 1b, c). Transgenic expression of the Cxcr4b–tFT reporter from a bacterial artificial chromosome (BAC) containing Cxcr4b cis-regulatory sequences (*cxcr4b:cxcr4b-tFT*) recapitulated Cxcr4b expression pattern and completely rescued the lateral line migration defect in Cxcr4b-deficient embryos (Extended Data Fig. 1), confirming that Cxcr4b–tFT was fully functional. Whole-mount embryo imaging revealed that Cxcr4b–tFT lifetime ratios varied in different embryonic tissues (Fig. 1d). A heat-shock-induced pulse of ubiquitous Cxcl12a (*hsp70:cxcl12a*) led to strongly enhanced receptor internalization and reduced lifetime ratios (Fig. 1e, f and Supplementary Video 1), with both the Cxcr4b–tFT reporter and endogenous receptor being sent for lysosomal degradation (Extended Data Fig. 2). Together, these experiments demonstrate that the Cxcr4b–tFT provides a dynamic readout of Cxcl12a chemokine activity in living embryos.

The Cxcr4b–tFT reporter was then used to study the pattern of chemokine activity across the lateral line primordium. Spatially resolved lifetime profiles along the tissue front–rear axis revealed a clear gradient in Cxcr4b activity, with twofold higher Cxcr4b–tFT lifetime ratios in follower cells than cells of the leading domain (Fig. 2a, c and Extended Data Fig. 3). Consistently, anti-Cxcr4b immunostainings confirmed that endogenous receptor levels are reduced at the plasma membrane in leading cells (Extended Data Fig. 4). To rule out the possibility that the observed gradient in receptor age results from a higher transcription in the front, we measured lifetime ratios of a membrane-tethered version of the tFT reporter (mem-tFT) that does not respond to the chemokine. A gradient in lifetime ratio was non-detectable when mem-tFT was expressed under identical genetic control (Fig. 2b, d and Extended Data Fig. 5). We conclude that the Cxcr4b–tFT lifetime ratio gradient across the primordium results from graded chemokine-triggered receptor turnover. Furthermore, imaging of Cxcr4b–tFT lifetime ratios across the primordium during recovery from a pulse of uniform Cxcl12a chemokine expression revealed that the gradient in receptor turnover was re-established as the effect of the chemokine pulse subsided (Fig. 2e, f and Supplementary Video 2). These data are consistent with the proposal that the primordium can self-generate a front–rear chemokine activity gradient that it dynamically maintains through interaction with its environment.

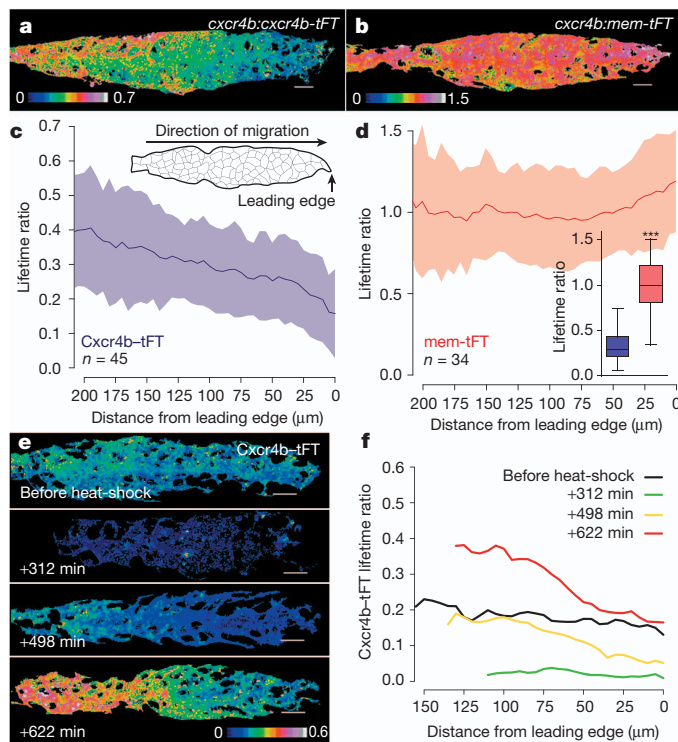
The establishment of a method to visualize the chemokine activity across the migrating tissue allowed us to determine the role of the endogenous chemokine receptors in shaping this gradient (Fig. 3a). Genetic inactivation of *cxcr4b* (mutant allele *cxcr4b*<sup>l26035</sup> (ref. 18)) decreased

<sup>1</sup>EMBL Heidelberg, Meyerhofstraße 1, 69117 Heidelberg, Germany. <sup>2</sup>Zentrum für Molekulare Biologie der Universität Heidelberg, Deutsches Krebsforschungszentrum, DKFZ-ZMBH Allianz, Im Neuenheimer Feld 282, 69120 Heidelberg, Germany. <sup>†</sup>Present addresses: CABD, Universidad Pablo de Olavide, Carretera de Utrera km1, 41013 Seville, Spain (A.F.-M.); MRC National Institute of Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK (G.V.).





Cxcr4b-tFT lifetime ratios (Fig. 3c–e), indicating that the endogenous receptor reduces Cxcl12a levels available to the Cxcr4b-tFT reporter. Lifetime ratios were more strongly reduced in *cxcr7* mutants (*cxcr7<sup>sa16</sup>* (ref. 19)), despite its expression being restricted to cells at the tissue rear (Fig. 3a–c). Conversely, knocking down *cxcl12a* in *cxcr7* mutants increased

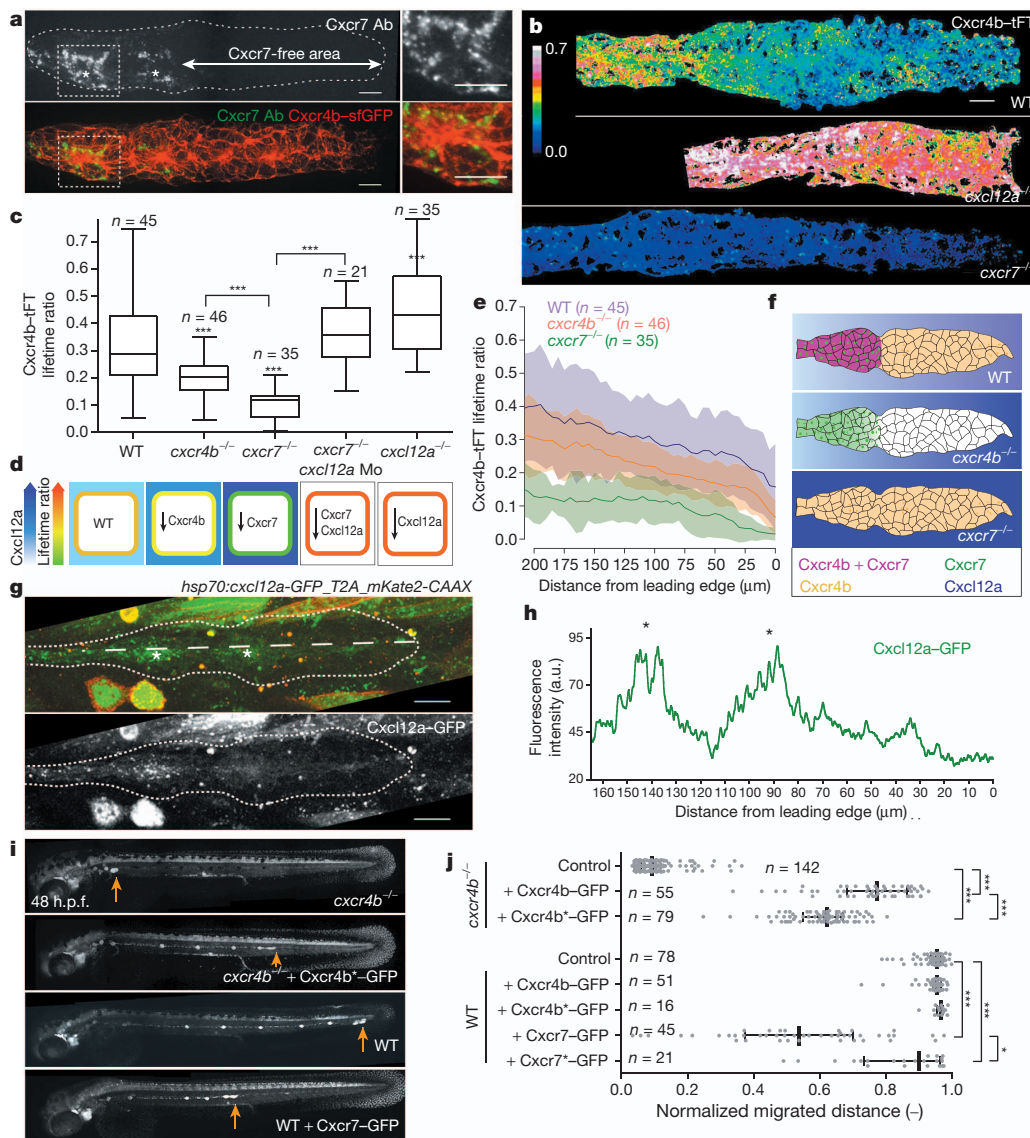


**Figure 1 | Tandem fluorescent protein timers (tFTs) allow quantification of chemokine-dependent receptor turnover *in vivo*.** **a**, Schema describing experimental rationale. Red/green intensity ratio increases as a function of protein age (left). Chemokine-dependent receptor-internalization rate determines ratio at plasma membrane (right). **b**, Top, schematic outline (left) and images of 7 h post-fertilization (h.p.f.) embryos injected with *cxcr4b-tFT* (middle) or *cxcr4b-tFT* plus *cxcl12a* mRNA (right), scale bar, 20  $\mu$ m. Bottom, ratiometric images of squared areas (calibration bar = normalized red/green ratio). **c**, Box-and-whiskers plot of Cxcr4b-tFT lifetime ratios as a function of *cxcl12a* mRNA concentration (one-way analysis of variance (ANOVA) test with Dunnett's multiple comparison procedure; control group: 0 ng  $\mu$ l<sup>-1</sup> *cxcl12a* mRNA; \*\**P* < 0.01). **d**, BAC *cxcr4b:cxcr4b-tFT* embryo at 32 h.p.f. Inset, magnification of dashed area (scale bars: embryo, 200  $\mu$ m; inset, 50  $\mu$ m). **e**, Cxcr4b-tFT response to heat-shock-induced expression of Cxcl12a (*hsp70:cxcl12a*, Supplementary Video 1). Scale bars: primordium, 20  $\mu$ m; insets, 5  $\mu$ m. Insets, magnification of boxed areas (brightness increased in bottom inset for visualization). **f**, Tissue-scale Cxcr4b-tFT lifetime ratio 5 h after heat-shock in *hsp70:cxcl12a* and controls (heat-shock = 45 min, box-and-whiskers plot, Welch's *t*-test, \*\*\**P* < 0.001). *n*, number of samples.

lifetime ratios to levels similar to *cxcl12a* null mutants (*cxcl12a<sup>l30516</sup>* (ref. 20), Fig. 3c), confirming that the effect of Cxcr7 on Cxcr4b turnover is through regulating chemokine availability. Spatially resolved profiles revealed a pronounced reduction and flattening of Cxcr4b-tFT lifetime ratios in the rear two-thirds of *cxcr7* mutant primordia (Fig. 3b, e). Interestingly, Cxcr4b-tFT lifetime ratio was also decreased in the leading domain that does not express Cxcr7, consistent with the idea that Cxcr7 regulates Cxcl12a availability across the entire primordium (Fig. 3e, f). Direct testing of the Cxcr7 sink model came from expressing a fluorescently tagged version of Cxcl12a (Cxcl12a-GFP) in cells adjacent to the migrating primordium. Cxcl12a-GFP accumulated in cells at the rear of the tissue, in Cxcr7-positive endosomes (Fig. 3g, h, Supplementary Video 3 and Extended Data Fig. 6). Thus, although the extracellular chemokine gradient generated by the primordium cannot be directly visualized using currently available tools, this pattern of increased accumulation of Cxcl12a-GFP at the tissue rear probably represents a 'complementary image' of its extracellular distribution. We therefore conclude that Cxcr7 sink function at the tissue rear patterns chemokine activity across the migrating primordium.

We next determined whether the activity of these receptors as sinks for Cxcl12a is of functional importance during collective migration. To this aim, we compared the ability of wild-type and internalization-defective versions of Cxcr4b (Cxcr4b\* (ref. 16)) and Cxcr7 (Cxcr7\*, Extended Data Fig. 7) to alter primordium migration. We expressed each receptor form using a primordium-specific Gal4 line that drives

**Figure 2 | Ligand-regulated chemokine receptor lifetime gradient indicates self-generated migration.** **a**, *cxcr4b:cxcr4b-tFT* (a) or a membrane-tethered version that does not respond to Cxcl12a (*cxcr4b:mem-tFT*) (b). **c**, *cxcr4b:cxcr4b-tFT* primordia at 32 h.p.f. Inset, tissue-scale lifetime ratios (box-and-whiskers plot, Welch's *t*-test, \*\*\**P* < 0.001). **d**, *cxcr4b:mem-tFT* primordia at 32 h.p.f. Inset, tissue-scale lifetime ratios (box-and-whiskers plot, Welch's *t*-test, \*\*\**P* < 0.001). **e**, Cxcr4b-tFT ratiometric images after pulse of Cxcl12a (*hsp70:cxcl12a*, Supplementary Video 2). **f**, Quantification of images shown in e. Scale bars, 10  $\mu$ m.



**Figure 3 | Role of Cxcr4b and Cxcr7 sink activities in chemokine gradient formation.** **a**, Expression pattern of Cxcr7 (anti-Cxcr7 antibody (Ab), green in bottom panel) and Cxcr4b (Cxcr4b-GFP, red in bottom panel). Cxcr7 vesicles cluster in apically constricted cells of organ progenitors (asterisks, insets). **b**, Ratiometric images of Cxcr4b-tFT primordia in wild type (WT, top), *cxcl12a*<sup>-/-</sup> (middle) and *cxcr7*<sup>-/-</sup> (bottom). Calibration bar shows the normalized red/green ratio. **c**, Box-and-whiskers plot of tissue-scale Cxcr4b-tFT lifetime ratios at 32 h.p.f. (Welch's *t*-test; asterisks refer to comparison with WT if not specified otherwise; \*\*\**P* < 0.001). *cxcl12a* Mo, *cxcl12a* morpholino. **d**, Schematic interpretation of data in **c**. **e**, Spatially resolved Cxcr4b-tFT ratio profiles in WT (blue), *cxcr4b*<sup>-/-</sup> (orange) and *cxcr7*<sup>-/-</sup> (green). Line, median; shaded area, median absolute deviation (MAD). **f**, Schematic interpretation of data in **e**. **g**, **h**, Cxcl12a-GFP expressed outside primordium is internalized at tissue rear. **g**, Mosaic expression of *hsp70:cxcl12a-GFP\_T2A\_mKate2-CAAX* showing Cxcl12a-GFP distribution (green) and secreting cells (red) 1 h after heat-shock (maximum intensity projection, stack in Supplementary Video 3). **h**, Mean Cxcl12a-GFP intensities (along dashed line in **g**). **i**, Primordium migration (arrows) after Gal4/UAS expression of indicated receptor. **j**, Dot-plot, median and interquartile range of normalized distance migrated (one-way ANOVA test with Bonferroni's multiple comparison adjustment, \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001). Scale bars, 10 μm (20 μm in **g**).

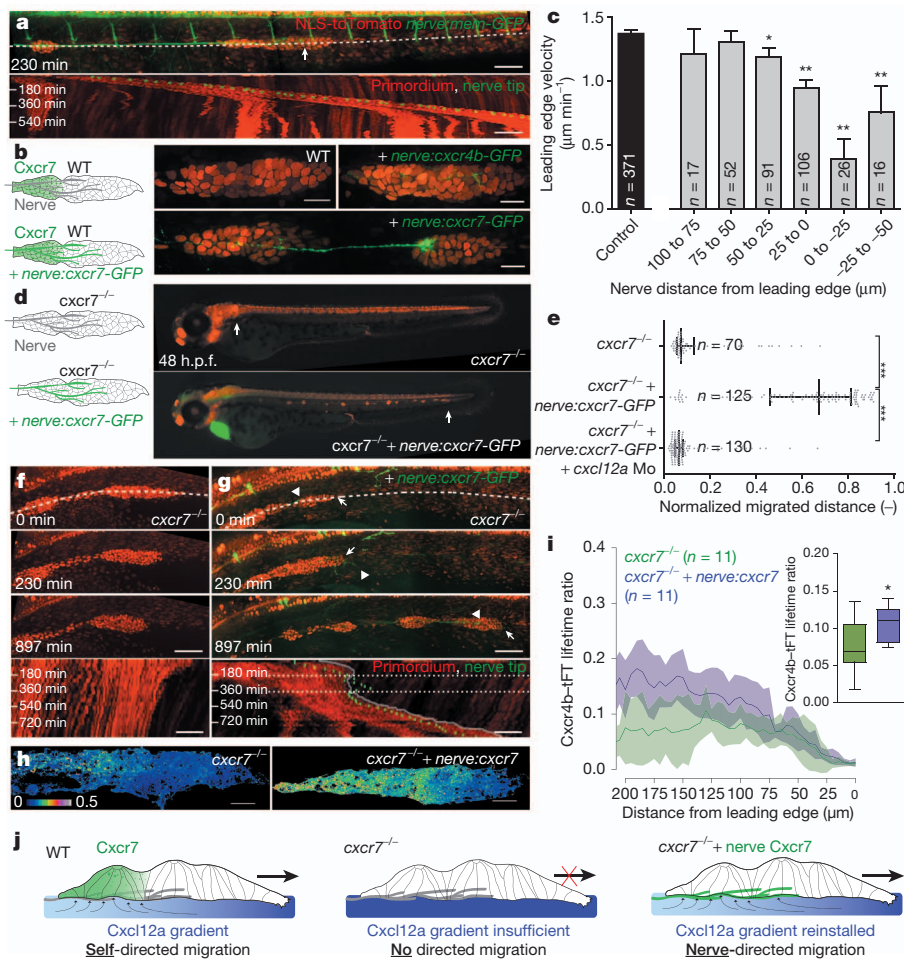
high-level expression in a mosaic pattern across the collective<sup>21</sup>. Overexpression of wild-type Cxcr4b-GFP had no influence on primordium migration and was able to rescue *cxcr4b* mutant primordia (Fig. 3j). Overexpression of Cxcr4b\*-GFP gave similar results in both assays (Fig. 3i, j), demonstrating that Cxcr4b sink activity is dispensable for directional collective migration. By contrast, overexpression of wild-type Cxcr7-GFP markedly impaired migration (Fig. 3i, j) and caused elongation of the tissue (Extended Data Fig. 8), as expected from expanding the expression of a Cxcl12a sink within the primordium. By comparison, overexpression of Cxcr7\*-GFP showed strongly reduced ability to interfere with migration (Fig. 3j), indicating that internalization is necessary for Cxcr7 activity in this context.

To test whether Cxcr7 sink activity is sufficient for migration, as predicted by the self-generated gradient model, we engineered an external source of Cxcr7 that followed the collective, achieved by expressing the receptor in the posterior lateral line nerve (pLLN) that extends under the primordium en route without influencing its movement<sup>22</sup> (Fig. 4a and Supplementary Video 4). Nerve expression of Cxcr7-GFP, achieved using a neural-specific promoter<sup>23</sup> (*nerve:cxcr7-GFP*) proved sufficient to interfere with the migration and morphology of wild-type primordia, with speed being reduced and the primordium becoming

elongated (Fig. 4b, c and Supplementary Video 5). Interestingly, migration decelerated when the Cxcr7-expressing nerve advanced into the leading domain (Fig. 4c), indicating that the position of the Cxcr7 sink dynamically influences tissue behaviour. Moreover, *nerve:cxcr7-GFP* expression reinstalled directional migration in *cxcr7* mutant primordia with remarkable efficiency (Fig. 4d, e). Time-lapse imaging of these nerve-guided tissues again highlighted the importance of Cxcr7 positioning, with directional migration arresting when the nerve advanced into the leading domain and recommencing when it retracted to trailing regions (Fig. 4f, g and Supplementary Videos 6–8). Cxcr4b-tFT lifetime ratio analysis revealed a chemokine activity gradient across primordia in *nerve:cxcr7*-rescued embryos when compared to *cxcr7* mutant siblings, confirming that extrinsic Cxcr7 sink activity is sufficient to promote both directional migration and self-generated gradient formation (Fig. 4h–j).

The data presented here provide compelling experimental support for the idea that migrating collectives can determine their own directionality via a self-generated gradient mechanism. This has a number of implications for our understanding of cell movement in complex systems. Most obviously, autonomous dynamic remodelling of extracellular guidance cues by migrating tissues circumvents the necessity





**Figure 4 | Chemokine gradient formation by Cxcr7 drives collective migration.** **a**, Still image of primordium (*cxcr4b:NLS-tdTomato*, red) and nerve (*nerve:mem-GFP*, green) at time-point 230 min of Supplementary Video 4, with kymograph (below). Arrow indicates nerve tip. **b**, Nerve-specific expression of Cxcr7-GFP (bottom), but not Cxcr4b-GFP (top right), alters primordium morphology (*cxcr4b:NLS-tdTomato*, red). **c**, Primordium leading-edge velocities as function of nerve distance (mean + s.e.m., one-way ANOVA test with Dunnett's multiple comparison procedure; \* $P < 0.05$ , \*\* $P < 0.01$ ). **d**, Cxcl12a-dependent primordium migration rescue by nerve-specific expression of Cxcr7-GFP. **d**, Primordium migration (arrows) in *cxcr7*<sup>-/-</sup> (top), plus *nerve:cxcr7-GFP* (bottom). **e**, Dot-plot, median and interquartile range of normalized distance migrated (one-way ANOVA test with Bonferroni's multiple comparison adjustment, \*\*\* $P < 0.001$ ). **f, g**, Time-lapse images of *cxcr7*<sup>-/-</sup> (f, Supplementary Video 6) and *cxcr7*<sup>-/-</sup> plus *nerve:cxcr7-GFP* (g, Supplementary Video 7) and corresponding kymographs (below). White dotted lines indicate nerve ahead of primordium. **h**, Ratiometric images of *cxcr7*<sup>-/-</sup> (left) and *cxcr7*<sup>-/-</sup> plus *nerve:cxcr7* primordia (right) at 28 h.p.f. **i**, Spatially resolved lifetime profiles of *cxcr7*<sup>-/-</sup> (green) and *cxcr7*<sup>-/-</sup> plus *nerve:cxcr7* (blue) primordia at 28 h.p.f. Inset, tissue-scale lifetime ratios (box-and-whiskers plot, Welch's *t*-test; \* $P = 0.012$ ). **j**, Model for nerve-rescue experiment. Scale bars, 20  $\mu$ m (b, h), 50  $\mu$ m (a, f, g).

for pre-existing long-range gradients, which may have spatial limitations. Moreover, it could coordinate the migration of multiple collectives within organisms, where gradients generated by one moving tissue accelerate or retard the migration of another. Our finding that the relocation of Cxcr7 allows the trailing nerve to direct the tissue that normally guides it shows that such higher-level coordination is possible. These experiments further show that signalling functions assigned to Cxcr7, such as  $\beta$ -arrestin recruitment<sup>15</sup>, regulation of Cxcr4 expression<sup>4</sup> or heterodimerization with Cxcr4 (ref. 24), are dispensable for this process, thus providing further support for signalling-independent activities for Cxcr7 *in vivo*<sup>12</sup> and encouraging us to re-evaluate its role in lateral line migration<sup>20</sup> (Supplementary Discussion and Extended Data Fig. 9). Here we have used protein lifetime reporters to reveal an autonomously generated chemokine activity gradient across the axis of a migrating collective. Although a number of cellular mechanisms could influence receptor turnover rates across tissues<sup>25</sup>, we demonstrate that Cxcr7 sink activity is required for and sufficient to pattern chemokine activity across the primordium. In the future it will be interesting to study how the tissue-scale chemokine gradient is converted to directed collective migration. For example, whether the self-generated gradient acts at the single-cell level, by directing the migration of individual cells within the leading domain, remains an open question. An attractive alternative, albeit non-mutually exclusive, proposal is that the function of Cxcr7 is to polarize motility at the tissue-scale by reducing chemokine activity levels to below a threshold, thus allowing follower cells behind the leading domain to adopt a different migratory behaviour. Finally, the generic nature of the lifetime imaging approach described here should facilitate the analysis of guidance signalling and self-generated gradient formation in other *in vivo* contexts. Given that Cxcr4 (ref. 26) and Cxcr7 (refs 27, 28) have important roles in cancer,

it is tempting to speculate that a mechanism of this kind could promote the self-directed migration of tumour cells<sup>6</sup>.

## METHODS SUMMARY

**Generation of Cxcr4b-tFT lifetime reporter BACs.** Cxcr4b was tagged at the carboxy terminus with a TagRFP-sfGFP tFT through modification of the BAC clone CH211-145M5, containing 250 kilobases of genomic sequence, using standard ET recombinering technology<sup>29</sup> (Gene-Bridges).

**Imaging of tFT reporter lines.** Embryos were mounted in 1% low melting point (LMP) agarose in glass-bottom dishes (MatTek Corporation) and imaged on a Zeiss LSM 780 confocal microscope. High-resolution z-stacks of primordia were captured using a 63 $\times$  water immersion objective (C-APOCHROMAT, 1.2 numerical aperture, Zeiss). Green and red signals were acquired sequentially using Gallium-Arsenide-Phosphide (GaAsP) detectors. A reference solution of purified mCherry-sfGFP fusion protein<sup>17</sup> was used to normalize for intensity fluctuations affecting the ratio. Low-resolution z-stacks of the entire embryo were acquired with a 20 $\times$  objective (EL Plan-NEOFLUAR, 0.5 numerical aperture, Zeiss). Red and green images correspond to merged Gaussian filtered and flattened (sum projection) dual-colour z-stacks, whereas ratiometric images were generated from sum projections of segmented z-stacks. Colour-coding, if not specified otherwise, indicates red/green ratio after normalization by the reference solution.

**Image analysis of lifetime ratios.** Custom software written in R was used for automated data analysis, using many features of the open source R/Bioconductor image analysis package EBImage<sup>30</sup>. The complete image analysis code is available as part of the Bioconductor data package DonaPLLP2013 (<http://www.bioconductor.org>). **Statistical analysis.** Statistical analysis of tissue-scale lifetime ratios was performed in R (Extended Data Fig. 10). The data set and vignette required to reproduce this analysis with runnable R code are available as part of the Bioconductor data package DonaPLLP2013 (<http://www.bioconductor.org>). GraphPad Prism 4.0 (GraphPad Software) was used for all other statistical analysis. Details of statistical tests are indicated in the figure legends. In box-and-whiskers plots the middle bar is the median, the hinges show the interquartile range (IQR) and the notches extend to  $\pm 1.58 \text{ IQR}/\sqrt{n}$ , where  $n$  is the number of samples.



**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 8 March; accepted 5 September 2013.**

**Published online 25 September 2013.**

1. Friedl, P. & Gilmour, D. Collective cell migration in morphogenesis, regeneration and cancer. *Nature Rev. Mol. Cell Biol.* **10**, 445–457 (2009).
2. Rørth, P. Whence directionality: guidance mechanisms in solitary and collective cell migration. *Dev. Cell* **20**, 9–18 (2011).
3. Montell, D. J. Morphogenetic cell movements: diversity from modular mechanical properties. *Science* **322**, 1502–1505 (2008).
4. Dambly-Chaudière, C., Cubedo, N. & Ghysen, A. Control of cell migration in the development of the posterior lateral line: antagonistic interactions between the chemokine receptors CXCR4 and CXCR7/RDC1. *BMC Dev. Biol.* **7**, 23 (2007).
5. Scherber, C. *et al.* Epithelial cell guidance by self-generated EGF gradients. *Integr. Biol.* **4**, 259–269 (2012).
6. Friedl, P., Locker, J., Sahai, E. & Segall, J. E. Classifying collective cancer cell invasion. *Nature Cell Biol.* **14**, 777–783 (2012).
7. Ghysen, A. & Dambly-Chaudière, C. The lateral line microcosmos. *Genes Dev.* **21**, 2118–2130 (2007).
8. David, N. B. *et al.* Molecular basis of cell migration in the fish lateral line: role of the chemokine receptor CXCR4 and of its ligand, SDF1. *Proc. Natl Acad. Sci. USA* **99**, 16297–16302 (2002).
9. Haas, P. & Gilmour, D. Chemokine signaling mediates self-organizing tissue migration in the zebrafish lateral line. *Dev. Cell* **10**, 673–680 (2006).
10. Streichan, S. J., Valentin, G., Gilmour, D. & Hufnagel, L. Collective cell migration guided by dynamically maintained gradients. *Phys. Biol.* **8**, 045004 (2011).
11. Weijer, C. J. Collective cell migration in development. *J. Cell Sci.* **122**, 3215–3223 (2009).
12. Boldajipour, B. *et al.* Control of chemokine-guided cell migration by ligand sequestration. *Cell* **132**, 463–473 (2008).
13. Naumann, U. *et al.* CXCR7 functions as a scavenger for CXCL12 and CXCL11. *PLoS ONE* **5**, e9175 (2010).
14. Lee, E. *et al.* CXCR7 mediates SDF1-induced melanocyte migration. *Pigment Cell Melanoma Res.* **26**, 58–66 (2013).
15. Rajagopal, S. *et al.*  $\beta$ -arrestin- but not G protein-mediated signaling by the 'decoy' receptor CXCR7. *Proc. Natl Acad. Sci. USA* **107**, 628–632 (2010).
16. Minina, S., Reichman-Fried, M. & Raz, E. Control of receptor internalization, signaling level, and precise arrival at the target in guided cell migration. *Curr. Biol.* **17**, 1164–1172 (2007).
17. Khmelinskii, A. *et al.* Tandem fluorescent protein timers for *in vivo* analysis of protein dynamics. *Nature Biotechnol.* **30**, 708–714 (2012).
18. Knaut, H., Werz, C. & Geisler, R. The Tübingen 2000 Screen Consortium & Nüsslein-Volhard, C. A zebrafish homologue of the chemokine receptor Cxcr4 is a germ-cell guidance receptor. *Nature* **421**, 279–282 (2003).
19. Kettleborough, R. N. W. *et al.* A systematic genome-wide analysis of zebrafish protein-coding gene function. *Nature* **496**, 494–497 (2013).
20. Valentin, G., Haas, P. & Gilmour, D. The chemokine SDF1a coordinates tissue migration through the spatially restricted activation of Cxcr7 and Cxcr4b. *Curr. Biol.* **17**, 1026–1031 (2007).
21. Distel, M., Wullmann, M. F. & Köster, R. W. Optimized Gal4 genetics for permanent gene expression mapping in zebrafish. *Proc. Natl Acad. Sci. USA* **106**, 13365–13370 (2009).
22. Gilmour, D., Knaut, H., Maischein, H.-M. & Nüsslein-Volhard, C. Towing of sensory axons by their migrating target cells *in vivo*. *Nature Neurosci.* **7**, 491–492 (2004).
23. Peri, F. & Nüsslein-Volhard, C. Live imaging of neuronal degradation by microglia reveals a role for vO-ATPase a1 in phagosomal fusion *in vivo*. *Cell* **133**, 916–927 (2008).
24. Levoye, A., Balabanian, K., Baleux, F., Bachelier, F. & Lagane, B. CXCR7 heterodimerizes with CXCR4 and regulates CXCL12-mediated G protein signaling. *Blood* **113**, 6085–6093 (2009).
25. Nakayama, M. *et al.* Spatial regulation of VEGF receptor endocytosis in angiogenesis. *Nature Cell Biol.* **15**, 249–260 (2013).
26. Müller, A. *et al.* Involvement of chemokine receptors in breast cancer metastasis. *Nature* **410**, 50–56 (2001).
27. Hernandez, L., Magalhaes, M. A. O., Coniglio, S. J., Condeelis, J. S. & Segall, J. E. Opposing roles of CXCR4 and CXCR7 in breast cancer metastasis. *Breast Cancer Res.* **13**, R128 (2011).
28. Luker, K. E. *et al.* Scavenging of CXCL12 by CXCR7 promotes tumor growth and metastasis of CXCR4-positive breast cancer cells. *Oncogene* **31**, 4750–4758 (2012).
29. Zhang, Y., Muylers, J. P., Testa, G. & Stewart, A. F. DNA cloning by homologous recombination in *Escherichia coli*. *Nature Biotechnol.* **18**, 1314–1317 (2000).
30. Pau, G., Fuchs, F., Sklyar, O., Boutros, M. & Huber, W. EBIImage—an R package for image processing with applications to cellular phenotypes. *Bioinformatics* **26**, 979–981 (2010).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We are grateful to J. Ellenberg, F. Peri and S. De Renzis for comments on the manuscript, S. Streichan and the Gilmour laboratory for suggestions. We thank R. Koester and M. Distel for the lateral line Gal4-driver, E. Busch-Nentwich and D. Stemple for the *cxcr7* mutant identified by the Zebrafish Mutation Project, the European Molecular Biology Laboratory (EMBL) Monoclonal Antibody Facility for Cxcr7 and Cxcr4b antibodies, the EMBL Advanced Light Microscopy Facility for imaging assistance, B. Klaus from the EMBL Centre for Statistical Data Analysis for advice and A. Gruia for fish care. We acknowledge funding from the European Commission's FP7 Network of Excellence 'Systems Microscopy' (W.H. and J.D.B.), Marie Curie FP6 (A.F.-M.), European Molecular Biology Organization (A.Kh) and Deutsche Forschungsgemeinschaft SFB 488 (D.G.).

**Author Contributions** D.G. and E.D. designed study with input from M.K. and A.Kh. E.D. performed all experiments, C.Q. helped perform the nerve rescue experiments. J.D.B. and W.H. developed the data analysis methods. G.V. developed the BAC-complementation of Cxcr4b. A.Ku generated the *cxcr4b::NLS-tdTomato* line, S.D. generated the *UAS:cxcr7-GFP* line, L.R.N. designed and cloned the *hsp70:cxcl12a-GFP\_T2A\_mKate2-CAAX* construct, A.F.-M. designed and validated the Cxcr7 and Cxcr4b monoclonal antibodies. D.G. and E.D. wrote the paper with input from all authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.G. ([gilmour@embl.de](mailto:gilmour@embl.de)).

## METHODS

**Zebrafish stocks.** Zebrafish (*Danio rerio*) strains were maintained and staged as described previously<sup>31</sup>. Embryos were raised in E3 buffer at 30 °C, unless otherwise specified. The following mutant and transgenic strains were used: *cxc4b*<sup>26035</sup> (ref. 18), *cxc12a*<sup>130516</sup> (ref. 20), *cldnb:lyn-GFP*, *cxc7<sup>ma16</sup>* (ref. 19), *cxc4a<sup>um20</sup>* (ref. 32). *GAL4-UAS:mCherry* (ETL GA346) line was provided by R. Köster<sup>21</sup>. BAC transgenic lines are described below. Other transgenic lines were made using the Tol2kit<sup>33</sup>. All zebrafish experiments were conducted on embryos before 3 days post-fertilization, under the rules of the European Molecular Biology Laboratory and the guidelines of the European Commission, Directive 2010/63/EU.

**mRNA synthesis and plasmids list.** Capped mRNA was synthesized using mMessageMachine kit (Ambion). Used plasmids, generated as multisite-gateway clones (Invitrogen) using the Tol2kit<sup>33</sup>, are listed below (in brackets the combined entry clones are indicated). 6×UAS promoter was provided kindly by R. Köster. *Cxc4b* and *Cxc4b\** constructs were a gift from the E. Raz laboratory (*Cxc4b\** corresponds to construct 6 in ref. 16). *Cxc7\** was created by removing the C-terminal 40 amino acids from *Cxc7* complementary DNA<sup>20</sup>, following the strategy used in ref. 34 (Extended Data Fig. 7). Plasmids list is included here, where NBT = *Xenopus* neural-specific β-tubulin promoter<sup>23</sup> (corresponds to nerve promoter), mem = membrane tethering myristoylation sequence from Lyn kinase, T2A = viral 2A peptide, CAAX = CAAX prenylation sequence: SP6:*Cxc4b*-TagRFP-sfGFP (p5'E(SP6/CMV)/pME(*Cxc4b* noStop)/p3'E(TagRFP-sfGFP)), 6×UAS:*Cxc4b*-GFP (p5'E(6xUAS)/pME(*Cxc4b* noStop)/p3'E(GFP)), 6×UAS:*Cxc4b\**-GFP (p5'E(6xUAS)/pME(*Cxc4b\** noStop)/p3'E(GFP)), 6×UAS:*Cxc7*-GFP (p5'E(6xUAS)/pME(*Cxc7* noStop)/p3'E(GFP)), 6×UAS:*Cxc7\**-GFP (p5'E(6xUAS)/pME(*Cxc7\** noStop)/p3'E(GFP)), hsp70:*Cxc12a* (p5'E(hsp70)/pME(MCS)/p3'E(*Cxc12a*)), nerve:*Cxc7*-GFP (p5'E(NBT)/pME(*Cxc7* noStop)/p3'E(GFP)), nerve:*Cxc4b*-GFP (p5'E(NBT)/pME(*Cxc4b* noStop)/p3'E(GFP)), nerve:mem-GFP (p5'E(NBT)/pME(mem noStop)/p3'E(GFP)), nerve:*Cxc7* (p5'E(NBT)/pME(*Cxc7*)/p3'E(polyA)), hsp70:*Cxc12a*-GFP\_T2A\_mKate2-CAAX (p5'E(hsp70)/pME(*Cxc12a*-GFP)/p3'E(T2A\_mKate2-CAAX)).

**BAC transgenesis.** Four BACs were generated by modification of *Cxc4b* BAC clone CH211-145M5 using the ET recombineering system<sup>29</sup> (Gene Bridges). The following selectable targeting cassettes were used for recombination. Targeting cassettes: (1) Lyn-tdTomato/late SV40 poly(A) signal/FRT-Kan<sup>R</sup>-FRT; (2) TagRFP(noStop)-(GlyAla)<sub>5</sub>-sfGFP/late SV40 poly(A) signal/FRT-Kan<sup>R</sup>-FRT; (3) Lyn-TagRFP(noStop)-(GlyAla)<sub>5</sub>-sfGFP/late SV40 poly(A) signal/FRT-Kan<sup>R</sup>-FRT; (4) cry:CFP/Kan<sup>R</sup>; (5) GFP/late SV40 poly(A) signal/FRT-Kan<sup>R</sup>-FRT (in which cry = γ-crystallin promoter<sup>33</sup> and CFP = cyan fluorescent protein).

The following primers, encoding for left and right homology arms, were used to amplify the listed selectable targeting cassettes (indicated in brackets). The sequence homologous to the PCR template is underlined.

Forward primers (left homology arms): (1) 5'-CGGTGTCTCAGGGAACGGA GGACTATTTAAATGCTGCTGTGAGAGGAGGCGGTGACTGAAGTGTG GAGACTTATTGCGCTTTTGTAGCGAGCCTCTAGAGCCACCATG-3'; (2) 5'-GATGCTGACCAAGAAAAGGGGCGCTATATCATCTGTATCTACTGAA TCAGAGTCGTCCAGTGCAGTACGAGTATGGTGTCTAAGGGCGAAG-3'; (3) 5'-CGGTGTCTCAGGGAACGGAAGGACTATTTAAATGCTGCTGTGAGA GGAGGCGGTGTGACTGAAGTGTGAGACTTATTGCGCTTTTGTAGCA AGTTTGTACAAAAAGCAGGCT-3'; (4) 5'-CGCGGGGCATGACTATTGG CGCGCGGATCGATCCTTAATTAAGTCTACTAATTATGATCCTCTAGAT CAGATCT-3'; (5) 5'-TCCAGTCAACAAGATGCTGACCAAGAAAAGGGGGCC TATATCATCTGTATCTACTGAATCAGAGTGTCCAGTGCAGTACGAGT ATGGTGAGCAAGGGCGAGGAG-3'.

Reverse primers (right homology arms): (1, 5) 5'-CATGATACACTTAACTA GTCTTATTAATAACTTATCAATGGAATGTCTGATAACACAGCGTTAT AAATTAATCAACAAGCCAACTTACCTAAAGGGAACAAAAGCTGGGTACC-3'; (2) 5'-ACACAAAAATACTTTACAATGTACAAAACTGTAGTAAAGTCTCT GTTTTATAAGCTTAATCATCTGTTGGGACCCGTTGGCGGTATCT-3'; (3) 5'-CATGATACACTTAACTAGTCTTATTAATAACTTATCAATGG AATGTCTGATAACACAGCGTTATAAATTAATCACAAGCCAACTTACGC ACCCGTGGCGGTACTCT-3'; (4) 5'-GCGTAAGCGGGGCACATTTTCATTACC TCTTTCTCCGACCCGACATAGATATTACCCTGTTATCCCTAGAAACAG CTATGACCATGTAA-3'.

For *cxc4b:NLS-tdTomato*, cassette 1 was inserted after *cxc4b* first intron, leaving the first 5 amino acids encoded by the first exon, which when fused to the membrane-tethering myristoylation sequence from Lyn kinase gave rise to a nuclear localization sequence (NLS). *cxc4b:cxc4b-tFT* (receptor C-terminal fusion) and *cxc4b:mem-tFT* (receptor transcriptional reporter, a tFT tethered to the membrane by the mem sequence) were generated in two steps. First, cassettes 2 and 3 were inserted to replace *Cxc4b* STOP and ATG starting codons, respectively. Then, *cry:CFP* transgenic marker (cassette 4) was introduced into the pTARBAC2.1 backbone. For *cxc4b:cxc4b-GFP* (receptor C-terminal fusion)

BAC line *Cxc4b* STOP codon was replaced by cassette 5. FRT-flanked kanamycin-resistance cassette (FRT-Kan<sup>R</sup>-FRT) was removed with FLP recombinase in all modified BACs. The inserted fragments and the regions flanking the recombination sites were sequenced before purification of positive clones (Large Construct Kit, Qiagen). Embryos were injected at one-cell stage, raised to adulthood and screened for fluorescence expression.

***cxc12a* knockdown.** Previously described *cxc12a* morpholino MO3 (ref. 12) (5'-CTACTACGATCACTTTGAGATCCAT-3', Gene Tools) was diluted to 1 mM and injected at one-cell stage. Injection efficacy was monitored by scoring the number of arrested primordia in injected wild-type control embryos.

**Analysis of *Cxc4b*-tFT lifetime ratio in early embryos.** *cxc4b-tFT* mRNA (200 ng μl<sup>-1</sup>, drop size ~0.6 nl) was injected alone or in the presence of titrating concentrations of *cxc12a* mRNA (10, 25 and 50 ng μl<sup>-1</sup>) in one-cell stage embryos. Dechorionated 7 h.p.f. embryos were mounted in 0.8% low melting point (LMP) agarose and high-resolution confocal images were acquired as described in the section 'Imaging of tFT reporters'. After background subtraction, membrane fragments were manually defined using the freehand line tool in Fiji (line width = 0.9 μm) followed by calculation of the mean red and green fluorescence intensities. Red/green ratios were normalized by the ratio of the sfGFP-mCherry reference solution and averaged per embryo. One embryo represents one data point.

**Imaging of tFT reporters.** Embryos were mounted in 1% LMP agarose in glass-bottom dishes (MatTek Corporation) and imaged on a Zeiss LSM 780 confocal microscope at 30 °C. High-resolution 16-bit z-stacks of primordia (pixel size = 0.131 μm, z-space = 0.5 or 1 μm) were captured using a 63× water immersion objective (C-APOCHROMAT, 1.2 NA, Zeiss). Green and red signals (green: ex. 488 nm, em. 489–550 nm; red: ex. 561 nm, em. 568–664 nm, A.U. = 1) were acquired sequentially by line using Gallium-Arsenide-Phosphide (GaAsP) detectors. A reference solution of purified mCherry-sfGFP fusion protein<sup>17</sup> was used to normalize for intensity fluctuations affecting the ratio. Low-resolution 16-bit z-stacks of the entire embryo (z-space = 10 μm, pixel size = 0.83 μm) were acquired with a 20× objective (EL Plan-NEOFLUAR, 0.5 NA, Zeiss). Red and green images correspond to merged Gaussian filtered and flattened (sum projection) dual colour z-stacks, whereas ratiometric images were generated as described in the section 'Generation of ratiometric images'.

**Image analysis of lifetime ratios.** The complete image analysis code is available as part of the Bioconductor data package DonaPLLP2013 (<http://www.bioconductor.org>). Blind sample selection was not possible, as mutant phenotypes were easily identifiable; however, specimens of each genotype were randomly selected from age-matched pools and lifetime ratios were extracted using an automated analysis pipeline, preventing bias in assessing the outcome of each experiment. In brief, z-stacks of primordia were acquired as described above and corrected for background. The green channel was chosen as the reference channel owing to the fast maturation kinetics of sfGFP and, from this, a binary mask of the primordium membrane (membrane-mask) was obtained by smoothing and adaptive thresholding. Then, a binary tissue-mask was computed and the axis of migration, along which all samples were aligned, was extracted using principal component analysis. Only pixels lying within the intersection of membrane-mask and tissue-mask (tissue-membrane-mask) were considered for subsequent calculations. To reduce the effects of pixel noise and chromatic aberration, a three-dimensional running median across the tissue (cube side-length of 10 μm) was computed in both channels. Red/green fluorescence intensity ratios were then calculated and normalized using a mCherry-sfGFP<sup>17</sup> reference solution acquired with the same imaging conditions. This procedure was necessary to compensate for day-to-day fluctuations of the instrument and permitted direct comparison of samples imaged on different sessions. To obtain tissue-scale lifetime ratios we calculated the median of normalized red/green fluorescence intensity ratios within the tissue-membrane-mask. Alternatively, to obtain spatially resolved lifetime ratio profiles, the median ratio along the axis of migration *x* (that is, through *yz*) was computed.

**Generation of ratiometric images.** From the image analysis we obtained background subtracted, Gaussian smoothed and membrane-masked z-stacks of the primordium in both red and green channels, which were further processed with Fiji. Stacks were flattened using sum projection, red/green ratio images were created using the Image Calculator plugin. These were then further normalized by dividing by the ratio of the reference solution. For visualization, images were colour-coded using the ratio lookup table in Fiji.

**Statistical analysis.** Statistical analysis of lifetime ratios was performed using R. The data set and vignette required to reproduce the analysis with runnable R code are available as part of the Bioconductor data package DonaPLLP2013 (<http://www.bioconductor.org>). To generate the analysed data set, consisting of normalized tissue-scale lifetime ratios obtained as described in the section 'Image analysis of lifetime ratios', embryos were carefully staged and always imaged at the same developmental stage (32 h.p.f.) to ensure comparability. Moreover, a calibration solution of purified mCherry-sfGFP fusion protein<sup>17</sup> was used to normalize for

variability in imaging conditions. As the effect size was not known a priori, we ran initial samples to obtain a first estimate of it. From these experiments we estimated that approximately  $n > 20$  samples would be sufficient to obtain enough statistical power; however, we imaged sample sizes of up to  $n = 46$ , from at least three technical replicates per condition. Subsequent statistical testing demonstrated high statistical power owing to these large sample sizes. The distribution of all tissue-scale ratios is shown in Extended Data Fig. 10a, allowing inspection of the variation between samples. Consistency with assumptions of normal distribution of the data was assessed by generation of quantile-quantile (Q-Q) plots for each condition individually (Extended Data Fig. 10b). These indicated that the data were sufficiently close to being normally distributed, therefore two-sided Welch's *t*-tests for each of the comparisons of interest and multiple testing correction across these comparisons with Bonferroni's method were performed. An alternative, non-parametric test, the two-sided Mann-Whitney test (a two-sample Wilcoxon test), returned comparable results leading to the same biological interpretation of the data (Extended Data Fig. 10c).

GraphPad Prism 4.0 (GraphPad Software) was used for all other statistical analysis. Details of statistical tests are indicated in the figure legends. In box-and-whiskers plots the middle bar is the median, the hinges show the interquartile range (IQR) and the notches extend to  $\pm 1.58 \text{ IQR}/\sqrt{n}$ , where  $n$  is the number of samples.

**Heat-shock induction of Cxcl12a.** Embryos were raised and imaged at 28 °C to avoid leakage of the *hsp70* promoter and heat-shocked at 38 °C in a water bath for 30 min if not indicated otherwise. Imaging was performed as described in the section 'Imaging of tFT reporters'. For time-lapse experiments four tiles were acquired and the area corresponding to the primordium was manually defined for further analysis. A time interval of approximately 1 h was chosen to minimize photobleaching and allow multi-position acquisition.

**Mosaic expression of Cxcl12a-GFP.** Embryos were injected at one-cell stage with *hsp70:cxcl12a-GFP\_T2A\_mKate2-CAAX*, raised at 28 °C. 32 h.p.f. embryos were heat-shocked at 38 °C for 1 h, mounted as described and imaged on a Zeiss LSM 780 confocal microscope (63× objective, GaAsP detector, pixel size = 0.131 μm). Owing to low intensity of the GFP signal the pinhole was opened to 2 a.u. Stacks were flattened by maximum intensity projection for visualization and by sum projection to calculate mean fluorescence intensities across the primordium mid-line (line width = 7.2 μm).

**Immunostaining.** Whole-mount antibody staining was performed following standard procedures<sup>31</sup>. Monoclonal anti-Cxcr7 and anti-Cxcr4b (zebrafish) antibodies were generated by the EMBL Monoclonal Antibody Facility using the following peptide sequences as antigens: MSVNVNDFNDILDALGELNFC and MEFYDS IILDNSSDSGSGDYDGEELC, corresponding to the amino terminus of Cxcr7 and Cxcr4b receptors, respectively. Antibody specificity was confirmed by demonstrating a lack of immunostaining pattern in embryos mutant for the respective receptor. Rabbit-anti GFP (Torrey Pines Biolabs) was used for staining of Cxcl12a-GFP. Goat anti-mouse Alexa Fluor 647 and goat anti-rabbit Alexa Fluor 488 (Molecular Probes) were used as secondary antibodies. Imaging was performed on either a Zeiss LSM 780 confocal microscope as described or a PerkinElmer Impropvision Ultraview VoX Spinning disk confocal. In this case, a 63× water immersion objective (C-APOCHROMAT, 1.2 NA, Zeiss, pixel size = 0.106 μm, z-space = 0.5 μm) and the following imaging settings were used: sequential acquisition; dichroic: multi pass 405/488/568/640; green: ex. 488 nm, em. single BP 525(W50); far red: ex. 640 nm, em. double BP 485(W60), 705(W90). Images were smoothed by Gaussian filter and flattened using maximum intensity projection in Fiji.

**Time-lapse imaging and primordium migration analysis.** Time-lapse experiments for the analysis of primordium migration were carried out at 30 °C on a PerkinElmer Impropvision Ultraview VoX Spinning disk confocal using 25× multi-immersion objective (Plan-APOCHROMAT, 0.8 NA, Zeiss). Image analysis was performed in Fiji. z-stacks spanning the whole primordium (z-space = 1 μm) were

flattened by maximum intensity projection and stitched using the Pairwise Stitching plugin<sup>35</sup> where necessary. Kymographs were generated using the Kymograph plugin (line width = 2 μm). When GFP fluorescent nerve was present, the furthest axon growth cone was manually indicated with a point on the migration path before creating the kymograph.

Time-lapses for the analysis of primordium migration speed (Fig. 4c) were 3–5 h long (time interval = 10 min). Primordium and nerve tips were tracked using the Manual Tracking plugin. Primordium tip instantaneous velocities were extracted (directionality: plus = anterior to posterior, minus = posterior to anterior) and binned depending on the distance between the tip of Cxcr7-GFP-positive nerve and the leading edge of the primordium at that time-point (binning window = 25 μm). A minus sign was assigned to the distance when the nerve tip was in front of the primordium. Sixteen movies per condition (wild type or wild type plus nerve:Cxcr7) were analysed.

Overviews of 48 h.p.f. embryos were acquired using a Leica MZ16F fluorescence stereomicroscope. Distances from the ear to the primordium and from the ear to the tail of the fish were measured using Fiji and their ratio was calculated.

**LysoTracker Red DND-99 staining.** Dechorionated embryos were incubated for 90 min in 1 μM LysoTracker Red DND-99 dye (Molecular Probes) in fish E3 medium plus 1% dimethylsulphoxide (DMSO).

**Bafilomycin A1 Inhibition of lysosomal function.** Dechorionated embryos were incubated in E3 medium plus Bafilomycin A1 (Sigma) at a final concentration of 300 nM for 3 h or 3 μM for 40 min. Embryos were then either fixed in 4% paraformaldehyde or mounted in 1% LMP agarose and imaged immediately.

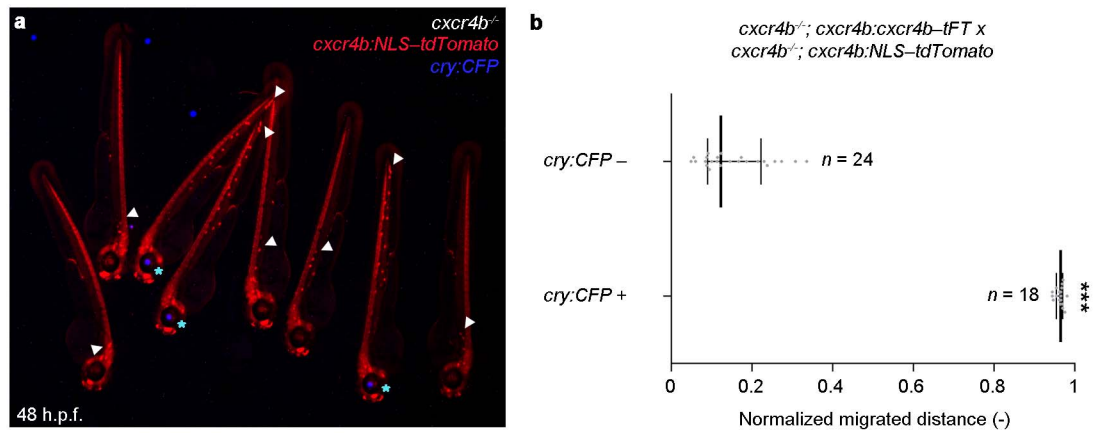
**Quantification of endogenous Cxcr4b degradation in response to Cxcl12a overexpression.** *cldnb:lyn-GFP/hsp70:cxcl12a* double-transgenic embryos were heat-shocked for 1 h at 38 °C and subsequently incubated for 3 h in the presence of Bafilomycin A1 at a final concentration of 300 nM or 0.01% DMSO (control). Control embryos (*cldnb:lyn-GFP* without *hsp70:cxcl12a*) were also heat-shocked. Embryos were then fixed in 4% paraformaldehyde and stained for Cxcr4b using the described monoclonal anti-Cxcr4b antibody. z-stacks spanning the whole primordium were acquired on a Zeiss LSM 780 confocal microscope. After background subtraction (rolling ball algorithm in Fiji) images were flattened using sum projection and Cxcr4b total fluorescence intensity was calculated within a primordium mask defined using the Lyn-GFP counter-label.

**Analysis of Cxcr4b distribution across the primordium.** *cldnb:lyn-GFP* transgenic embryos were fixed at 32 h.p.f. and stained for Cxcr4b using the aforementioned monoclonal anti-Cxcr4b antibody. Confocal stacks were acquired on a Zeiss LSM 780 as already described for tFT reporters imaging. The same pipeline used for the ratiometric analysis of Cxcr4b-tFT was followed, with two differences: the intersection of the green and red individual masks was used as membrane mask and profiles were normalized by mean ratio in the front 15 μm of the tissue.

**Cxcr4a's role in primordium migration.** Whole-mount *in situ* hybridization of Cxcr4a was performed using standard procedures. Cxcr4a probe and *cxcr4a*<sup>um20</sup> (ref. 32) were provided by A.F. Siekmann. *cxcr4b:NLS-tdTomato/cxcr4a*<sup>um20/+</sup> transgenic fish were in-crossed and primordium migration analysis was performed as already described.

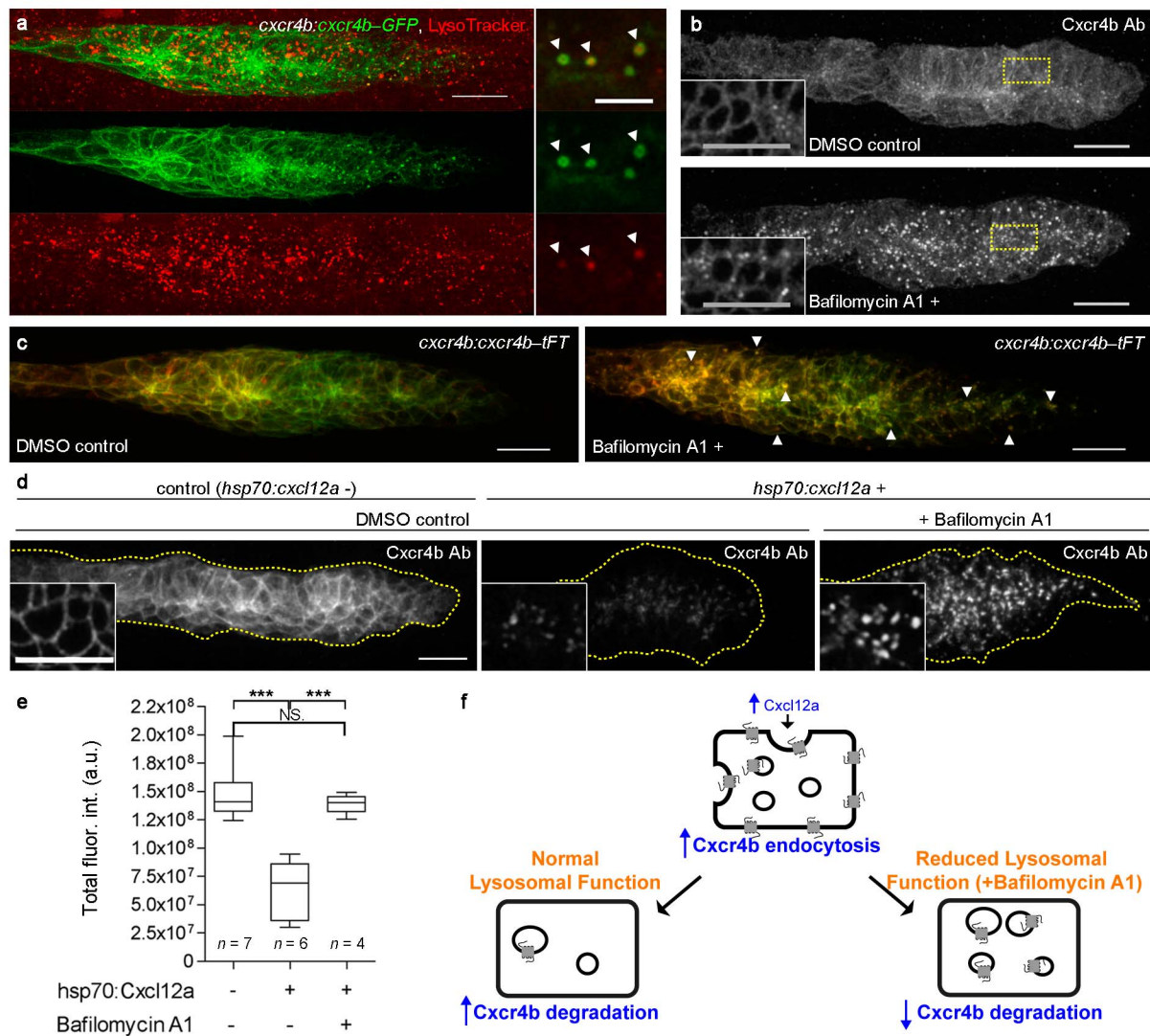
- Westerfield, M. *The Zebrafish Book: A Guide for the Laboratory use of Zebrafish (Brachydanio rerio)* 2nd edn (University of Oregon Press, 1993).
- Siekmann, A. F., Standley, C., Fogarty, K. E., Wolfe, S. A. & Lawson, N. D. Chemokine signaling guides regional patterning of the first embryonic artery. *Genes Dev.* **23**, 2272–2277 (2009).
- Kwan, K. M. *et al.* The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. *Dev. Dyn.* **236**, 3088–3099 (2007).
- Canals, M. *et al.* Ubiquitination of CXCR7 controls receptor trafficking. *PLoS ONE* **7**, e34192 (2012).
- Preibisch, S., Saalfeld, S. & Tomancak, P. Globally optimal stitching of tiled 3D microscopic image acquisitions. *Bioinformatics* **25**, 1463–1465 (2009).





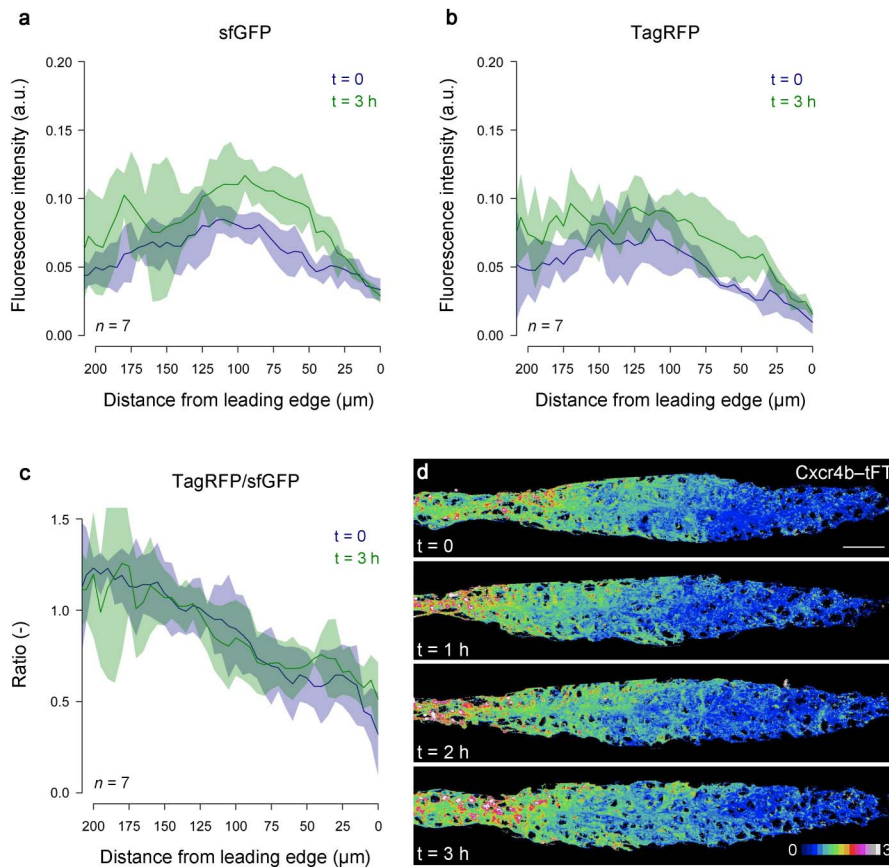
**Extended Data Figure 1 | *cxcr4b:cxcr4b-tFT* BAC rescues primordial migration defect in *cxcr4b* mutants.** **a**, Image of *cxcr4b* mutant embryos, where primordial migration (white arrowheads) is made visible by *cxcr4b:NLS-tdTomato* (red). Fish carrying *cxcr4b:cxcr4b-tFT* transgenes are identifiable by an independent transgenic marker that labels the lens blue

(*cry:CFP*, asterisks). **b**, In all *cxcr4b:cxcr4b-tFT* transgenic embryos the primordial migrated normally, whereas non-transgenics showed migration defects typical of *cxcr4b* mutants<sup>20</sup>. Dot-plot, median and interquartile range of the indicated number of samples are shown (Welch's *t*-test, \*\*\**P* < 0.0001).



**Extended Data Figure 2 | Cxcl12a-induced internalization targets Cxcr4b to lysosomal degradation.** **a–c**, Cxcr4b is sent for lysosomal degradation in wild-type embryos. **a**, *cxcr4b:cxcr4b-GFP* embryo stained with the lysosomal marker LysoTracker Red DND-99. Left, maximum intensity projection of a primordium confocal z-stack (scale bar, 20  $\mu$ m) showing Cxcr4b-GFP (green) and LysoTracker (red) localization. Right, magnification of single confocal plane showing LysoTracker and Cxcr4b-GFP co-localization (arrowheads; scale bar, 5  $\mu$ m). **b**, Impaired lysosomal acidification leads to accumulation of endogenous Cxcr4b in intracellular vesicles. Anti-Cxcr4b (maximum intensity projections) of Bafilomycin A1-treated embryos (bottom) and controls (top; DMSO (dimethylsulphoxide)). **c**, Bafilomycin A1 treatment results in an increased number of Cxcr4b-tFT vesicles. Images show maximum intensity

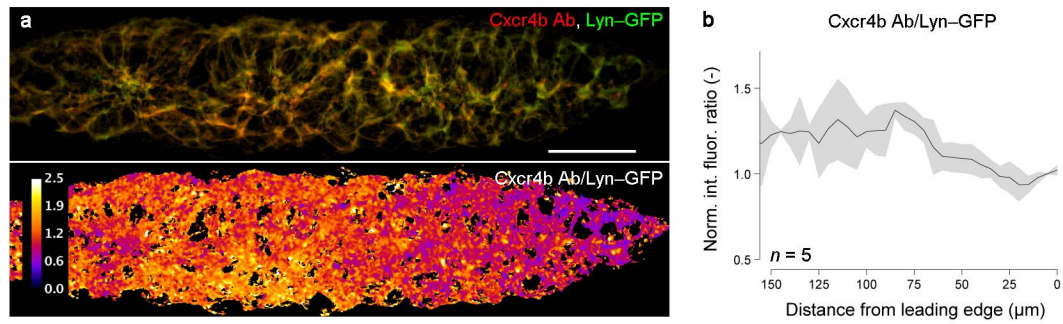
projections of z-stacks (scale bar, 20  $\mu$ m). **d–f**, High Cxcl12a levels cause increased degradation of endogenous Cxcr4b. **d**, Anti-Cxcr4b immunohistochemistry 3 h after heat-shock induction (1 h at 38  $^{\circ}$ C) of *hsp70:cxcl12a* (sum projection of confocal z-stacks). **e**, Quantification of Cxcr4b levels within the primordium (box-and-whiskers plot, one-way ANOVA test with Bonferroni's multiple comparison adjustment, \*\*\* $P$  < 0.001) showing significantly reduced levels in *hsp70:cxcl12a* transgenics compared to non-transgenic controls. This effect is suppressed by Bafilomycin A1 treatment, resulting in increased accumulation of Cxcr4b-positive intracellular vesicles (**f**). Insets show single plane magnification of the squared area. Scale bars, 20  $\mu$ m. NS, not significant.



**Extended Data Figure 3 | Cxcr4b-tFT lifetime ratio is stable over several hours.** **a–c**, Spatially resolved profiles of Cxcr4b-tFT-expressing primordia at 32 h.p.f. ( $t = 0$ ) and 35 h.p.f. ( $t = 3 \text{ h}$ ) show that although sfGFP (**a**) and TagRFP (**b**) median fluorescence intensities increase through time, their ratio instead stays constant and maintains a gradient across the front–rear axis of the

tissue (**c**). Median and MAD of 7 samples acquired on the same day with identical imaging conditions are shown. **d**, Ratiometric images of one representative sample through time. Lookup table = red/green fluorescence intensity ratio. Scale bar, 20  $\mu\text{m}$ . All lifetime measurements were performed during this time window, whenever possible.

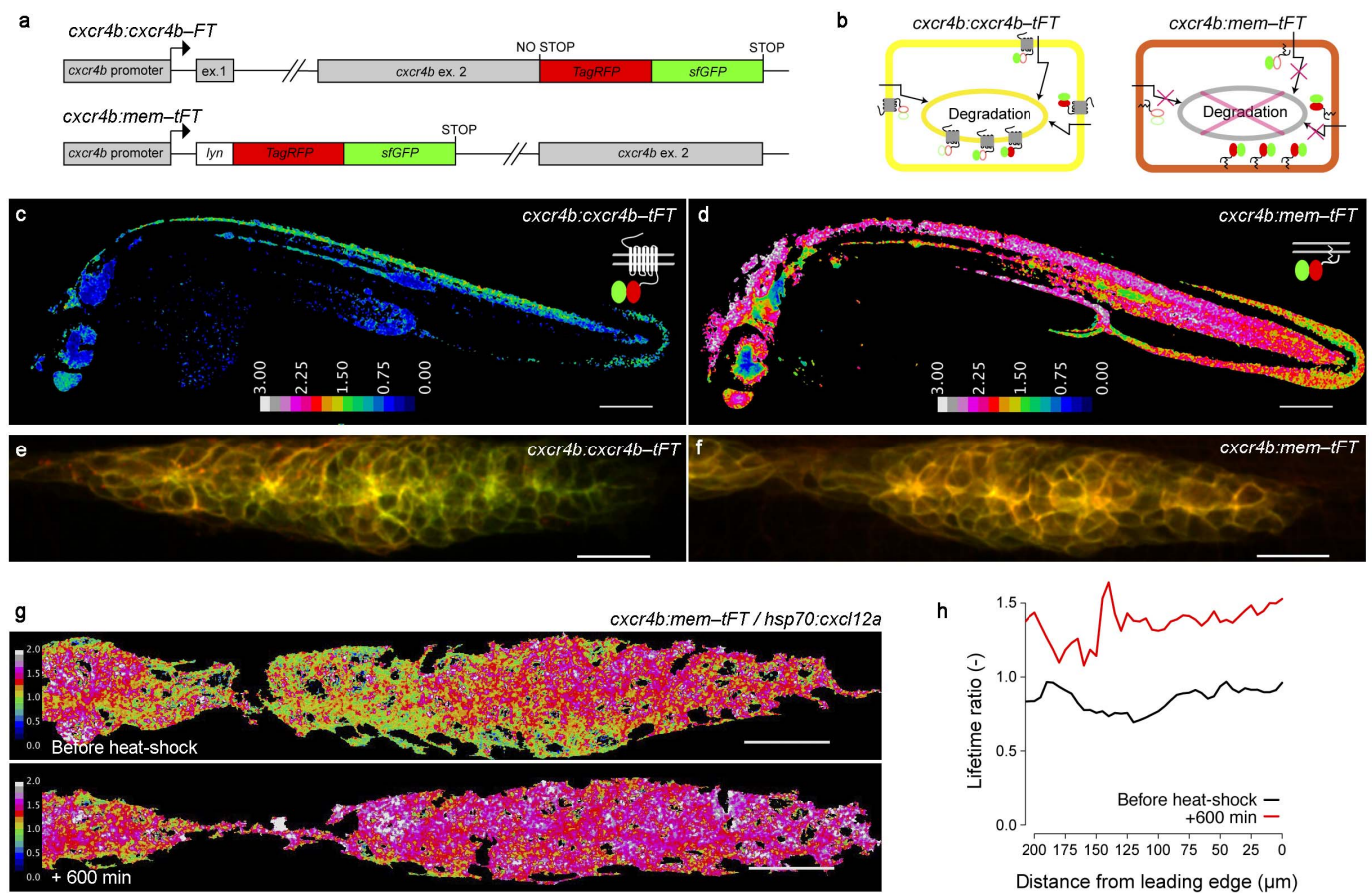




**Extended Data Figure 4 | Increased turnover in primordial leading domain results in graded distribution of endogenous Cxcr4b receptor.**

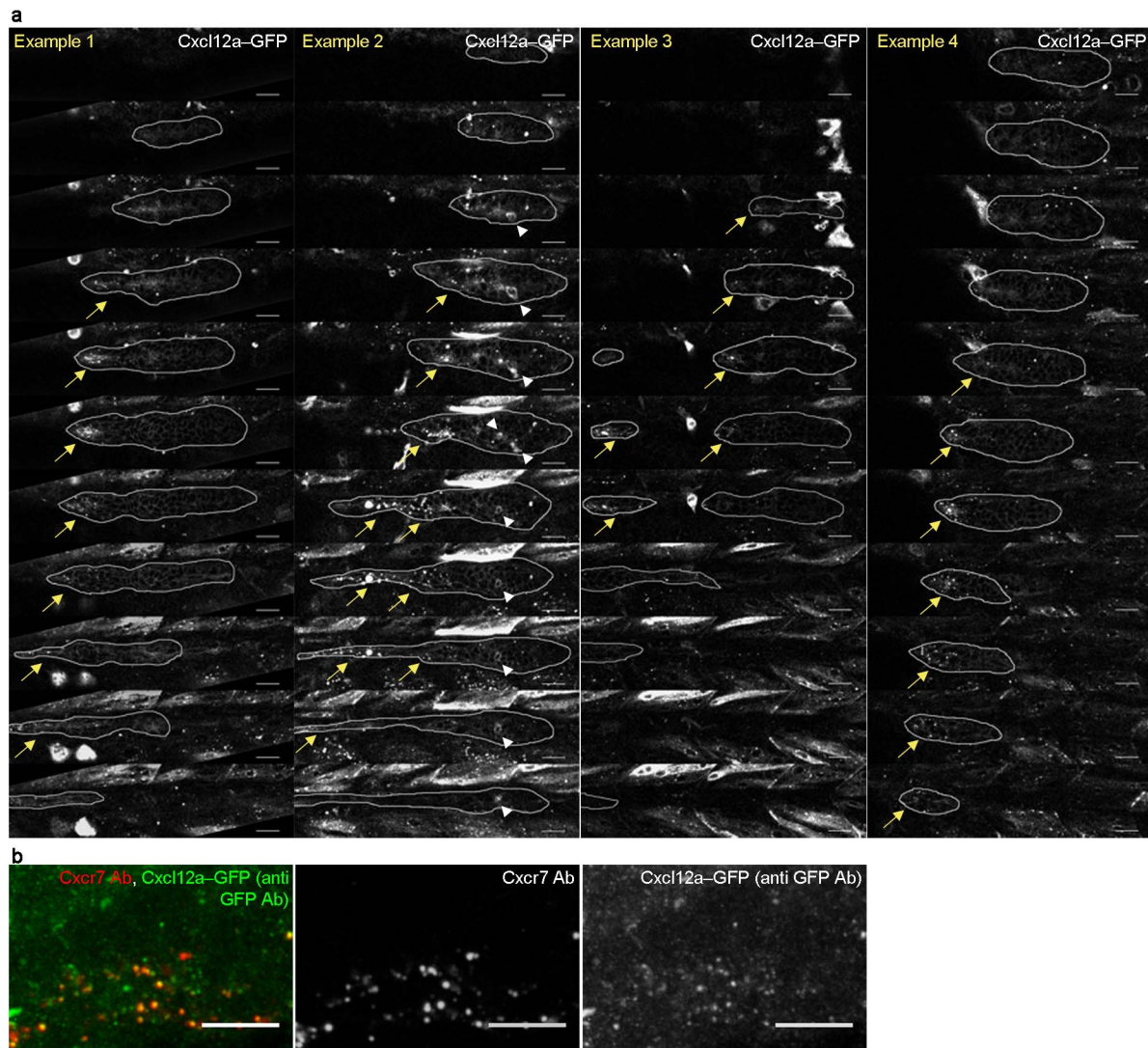
Anti-Cxcr4b immunohistochemistry reveals a gradient of endogenous receptor along the primordial. **a**, Cxcr4b antibody (red) and a homogeneous membrane counter-label given by *cldnb:lyn-GFP* transgenic line (green)

displayed in segmented (top, sum projection of segmented stack) and ratiometric form (bottom, calibration bar = Cxcr4b antibody/Lyn-GFP fluorescence intensity ratio). **b**, Cxcr4b antibody/Lyn-GFP fluorescence intensity ratio profiles. Median and MAD after normalization of individual profiles by the front-most 15 μm are plotted.



**Extended Data Figure 5 | Cxcr4b-tFT lifetime ratio gradient results from protein turnover not from a transcriptional gradient.** **a**, Schematic representation of the modified BACs used to create Cxcr4b-tFT and mem-tFT reporter lines. **b**, Predicted behaviour of Cxcr4b-tFT and mem-tFT in response to Cxcl12a. **c**, **d**, Ratiometric images of 36 h.p.f. Cxcr4b-tFT (**c**) and mem-tFT (**d**) of whole-mount transgenic embryos. Segmentation was performed using intensity-based thresholding on sfGFP channel. Scale bars, 200  $\mu$ m. Calibration

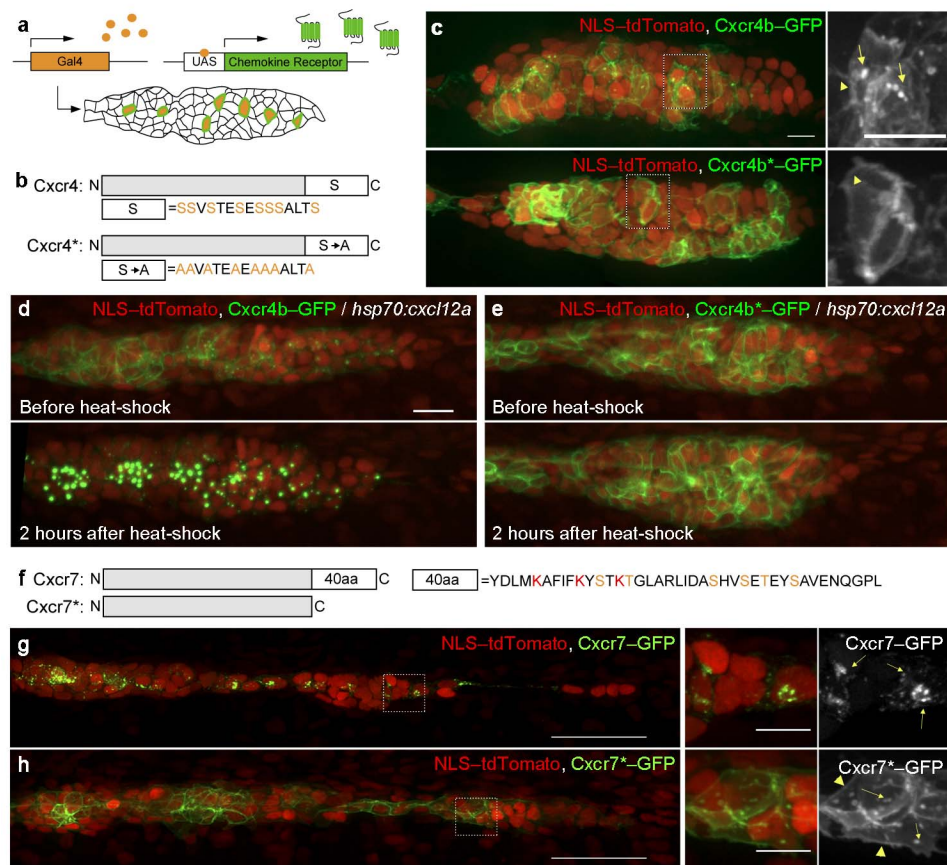
bar = normalized red/green ratio. **e**, **f**, Raw primordium images (merge of green and red channels) corresponding to ratiometric images shown in Fig. 2a, b. Scale bars, 20  $\mu$ m. **g**, Ratiometric images of *cxcr4b:mem-tFT* before (top) and 10 h after (bottom) heat-shock-induced pulse of Cxcl12a. Scale bars, 20  $\mu$ m. **h**, Spatially resolved lifetime ratio analysis of the sample in **g**. Although overall the ratio increases, during the long time lag, no front to rear gradient could be detected, in contrast to the behaviour of Cxcr4b-tFT (compare to Fig. 2e, f).



**Extended Data Figure 6 | Cxcr7-mediated internalization of Cxcl12a-GFP at the tissue rear.** **a**, Film-strip montages of z-stacks showing mosaic expression of Cxcl12a-GFP resulting in preferential internalization at the rear of the primordium (outlined by white line, corresponding to Supplementary Video 3). Stacks were acquired 1–2 h after heat-shock induction (1 h at 38 °C) of *hsp70:cxcl12a-GFP\_T2A\_mKate2-CAAX*-injected embryos. Accumulation

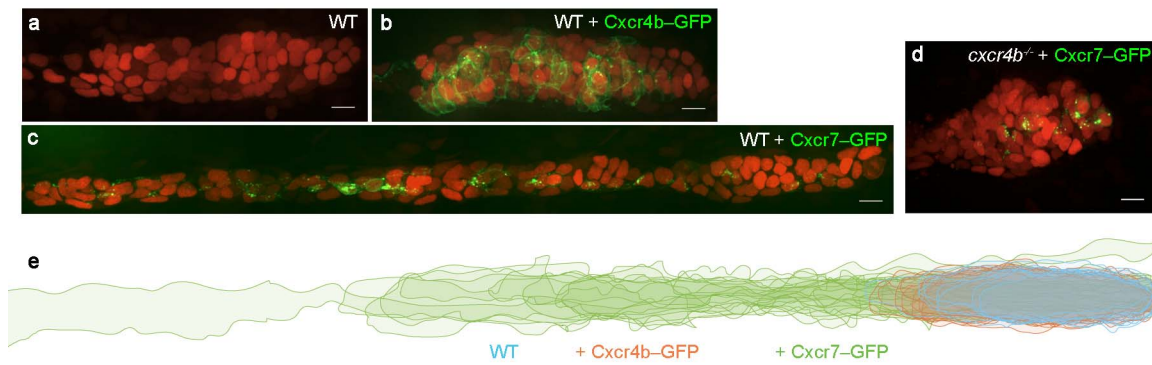
of Cxcl12a-GFP-positive vesicles in the primordium (yellow arrows) and Cxcl12a-GFP-expressing cells within the primordium are indicated (white arrowheads). Scale bar, 20 μm. **b**, Anti-GFP and anti-Cxcr7 immunohistochemistry showing co-localization of Cxcl12a-GFP (green) and Cxcr7 (red) in vesicles at the rear of the primordium. Images show maximum intensity projection. Scale bars, 10 μm.





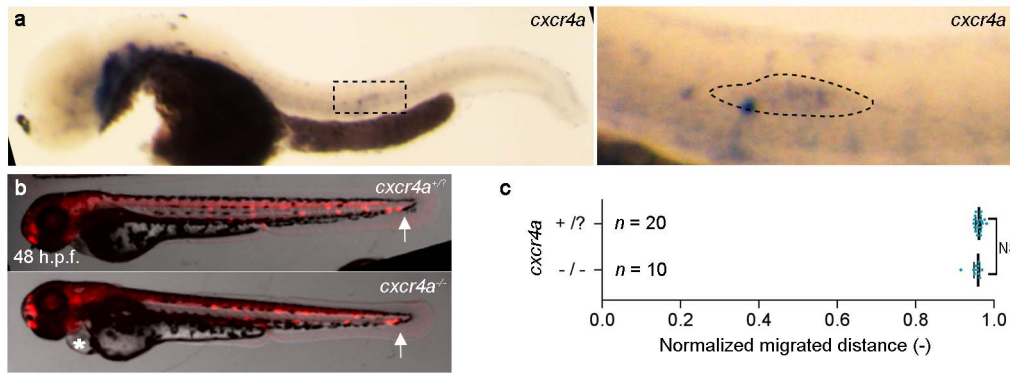
**Extended Data Figure 7 | Validation of internalization-defective chemokine receptor forms.** **a**, Cartoon illustrating the Gal4/UAS strategy used to induce mosaic expression of chemokine receptors in the primordium. **b**, Schematic of internalization-defective Cxcr4b form (Cxcr4\*), as described in ref. 16. Mutated putative phosphorylation sites (S to A) are shown in orange. **c**, Examples of Gal4/UAS-driven expression of Cxcr4b-GFP (green or greyscale, top) or Cxcr4\*-GFP (green or greyscale, bottom) in *cxc12a:NLS-tdTomato* (red) transgenic embryos. Both receptor forms localize at the plasma membrane (arrowheads). Higher magnification of the squared area shows internalized Cxcr4b (arrows), but not Cxcr4\*, under unstimulated conditions, presumably in response to endogenous Cxcl12a. Scale bars, 10  $\mu$ m.

**d, e**, Heat-shock-induced Cxcl12a expression causes complete internalization of Cxcr4b-GFP (**d**), whereas Cxcr4b\*-GFP remains at the plasma membrane (**e**). Thus, Cxcr4b\* does not internalize Cxcl12a. **f**, An internalization-defective form of Cxcr7 (Cxcr7\*) was created by deleting the 40 C-terminal amino acids of the receptor. Putative ubiquitination and phosphorylation target residues, conserved in humans, are shown in red and orange, respectively. **g**, GFP fusion of full-length Cxcr7 localizes in vesicles (arrows in the inset), consistent with antibody stainings of endogenous Cxcr7 (Fig. 3a). **h**, C terminus deletion results in clear receptor re-localization to the plasma membrane (arrowheads). Cxcr7\* is only occasionally found in vesicles (arrows). Expression of both receptor forms was driven by Gal4/UAS. Scale bars, 50  $\mu$ m (10  $\mu$ m in the insets).



**Extended Data Figure 8 | Cxcr7-GFP misexpression alters migrating primordium morphology.** **a–c**, Primordia at 36 h.p.f. (marked by NLS-dtTomato) in wild type (WT) conditions (**a**) or misexpressing Cxcr4b-GFP (**b**) or Cxcr7-GFP (**c**) under the control of the Gal4/UAS system. **d**, *cxcr4b*<sup>-/-</sup> primordium misexpressing Cxcr7-GFP (green). Scale bars, 10  $\mu$ m. **e**, Overlay

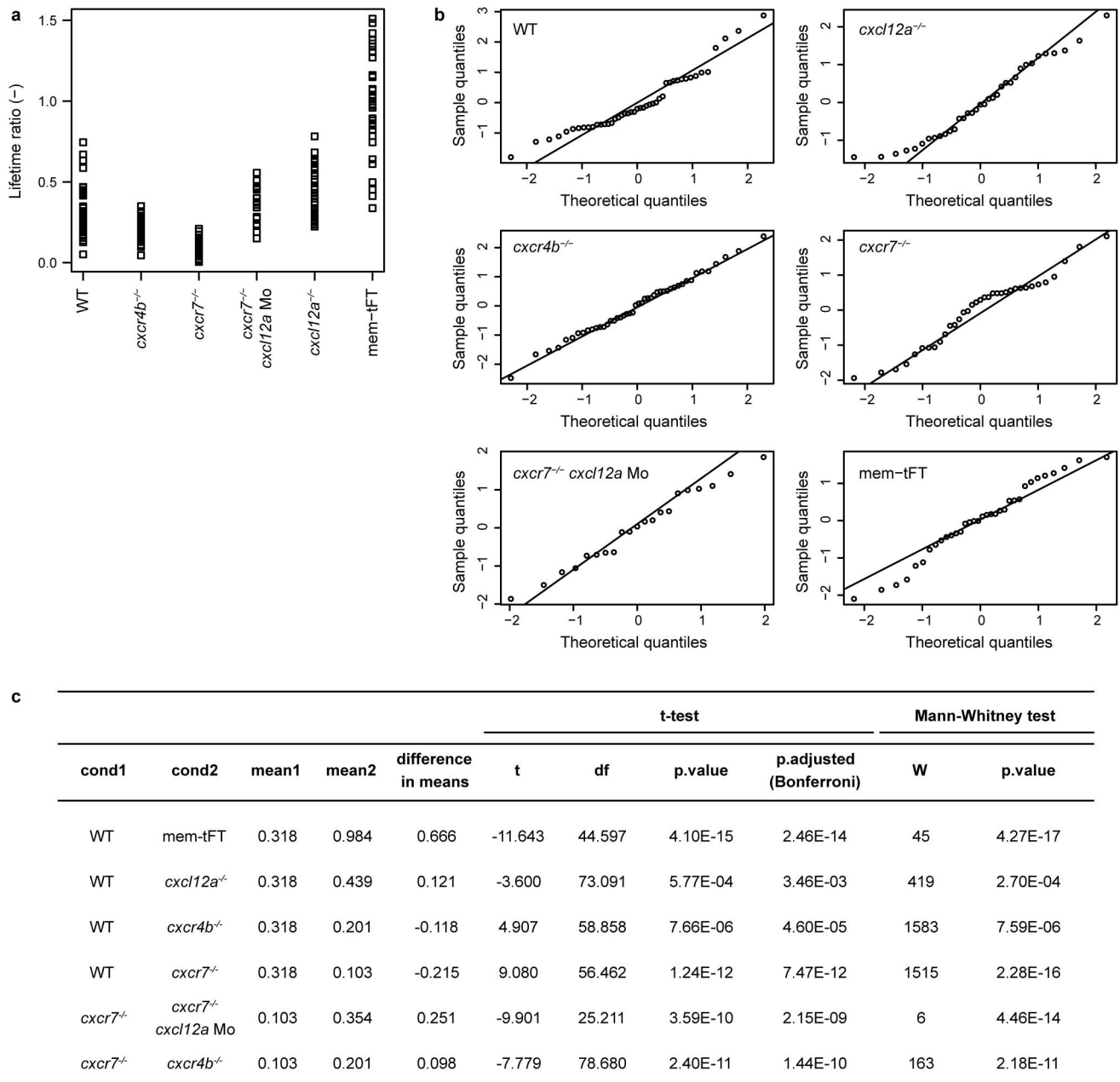
of outlines of several primordia in the same conditions as in **a–c**. Primordia are aligned by the leading edge (right). Ectopic Cxcr7-GFP, but not Cxcr4b-GFP, expression results in primordium elongation. This phenotype is Cxcr4b dependent, as elongated morphology is never observed in Cxcr4b mutants misexpressing Cxcr7-GFP (**d**).



**Extended Data Figure 9 | Cxcr4a is not required for primordium migration.** **a**, *In situ* hybridization reveals *cxcr4a* expression in the primordium (36 h.p.f. embryo). Inset shows higher magnification of the squared area comprising the primordium. **b**, Images show primordium migration (red = *cxcr4b:NLS-tdTomato*) at 48 h.p.f. in the presence (top) or absence (bottom) of Cxcr4a.

Asterisk marks heart oedema, characteristic of *cxcr4a*<sup>-/-</sup> (corresponding to *cxcr4a*<sup>um20/um20</sup>)<sup>32</sup>. **c**, Quantification of primordium migration in *cxcr4a*<sup>-/-</sup> embryos. The absence of Cxcr4a does not affect migration when Cxcr4b is present. Dot-plot, median, interquartile range and number of analysed samples are shown. Unpaired *t*-test.





**Extended Data Figure 10 | Statistical analysis of tissue-scale lifetime ratios.**  
**a**, Summary of tissue-scale lifetime ratio distributions at 32 h.p.f. (data

corresponding to Figs 2d and 3c). **b**, Quantile-quantile (Q-Q) plots of the analysed data sets. **c**, Statistical tests results.

# A high-resolution map of the three-dimensional chromatin interactome in human cells

Fulai Jin<sup>1\*</sup>, Yan Li<sup>1\*</sup>, Jesse R. Dixon<sup>1,2</sup>, Siddarth Selvaraj<sup>1,3</sup>, Zhen Ye<sup>1</sup>, Ah Young Lee<sup>1</sup>, Chia-An Yen<sup>1</sup>, Anthony D. Schmitt<sup>1,4</sup>, Celso A. Espinoza<sup>1</sup> & Bing Ren<sup>1,5</sup>

A large number of *cis*-regulatory sequences have been annotated in the human genome<sup>1,2</sup>, but defining their target genes remains a challenge<sup>3</sup>. One strategy is to identify the long-range looping interactions at these elements with the use of chromosome conformation capture (3C)-based techniques<sup>4</sup>. However, previous studies lack either the resolution or coverage to permit a whole-genome, unbiased view of chromatin interactions. Here we report a comprehensive chromatin interaction map generated in human fibroblasts using a genome-wide 3C analysis method (Hi-C)<sup>5</sup>. We determined over one million long-range chromatin interactions at 5–10-kb resolution, and uncovered general principles of chromatin organization at different types of genomic features. We also characterized the dynamics of promoter–enhancer contacts after TNF- $\alpha$  signalling in these cells. Unexpectedly, we found that TNF- $\alpha$ -responsive enhancers are already in contact with their target promoters before signalling. Such pre-existing chromatin looping, which also exists in other cell types with different extracellular signalling, is a strong predictor of gene induction. Our observations suggest that the three-dimensional chromatin landscape, once established in a particular cell type, is relatively stable and could influence the selection or activation of target genes by a ubiquitous transcription activator in a cell-specific manner.

We carried out Hi-C experiments to study the dynamic chromatin interactions in a primary human fibroblast cells (IMR90) in response to transient TNF- $\alpha$  signalling. Hi-C data from IMR90 cells before and after 1 h TNF- $\alpha$  treatment were combined, to produce a total of approximately 3.4 billion uniquely mapped paired-end reads from 6 biological replicates in each condition, among which approximately 1.4 billion are intra-chromosomal reads (Supplementary Tables 1 and 2). To accurately identify chromatin looping interactions with high sensitivity and resolution, we devised an improved data filtering strategy<sup>6</sup> based on the strand orientation of Hi-C paired-end reads (Supplementary Figs 1–6 and Supplementary Methods), which results in over 500 million high-confidence read pairs (Supplementary Tables 1 and 2), each representing a legitimate ligation event between two restriction fragments on the same chromosome. As we recognized that some reads may be due to random collision events between restriction fragments<sup>4,7</sup>, we also estimated the expected frequency between any two restriction fragments, and then fitted a negative binomial model to assess the significance of observed contact frequency (Supplementary Methods and Supplementary Figs 7–9). Compared to previous methods<sup>4</sup>, our data analysis method permits detection of chromatin interactions at short distance. For example, we observed asymmetric distribution of *cis*-contacts from highly expressed promoters to the immediate downstream gene bodies (Supplementary Fig. 10). This observation is reminiscent of a recent study showing interactions between a subset of exons and their promoters<sup>8</sup>. Interestingly, although such bias at promoters is correlated with elongation of RNA polymerase II, it

remains when transcription elongation is blocked by the positive transcription elongation factor b (P-TEFb) inhibitor flavopiridol (Supplementary Fig. 11), suggesting that the maintenance of promoter–gene-body contacts is independent of active transcription.

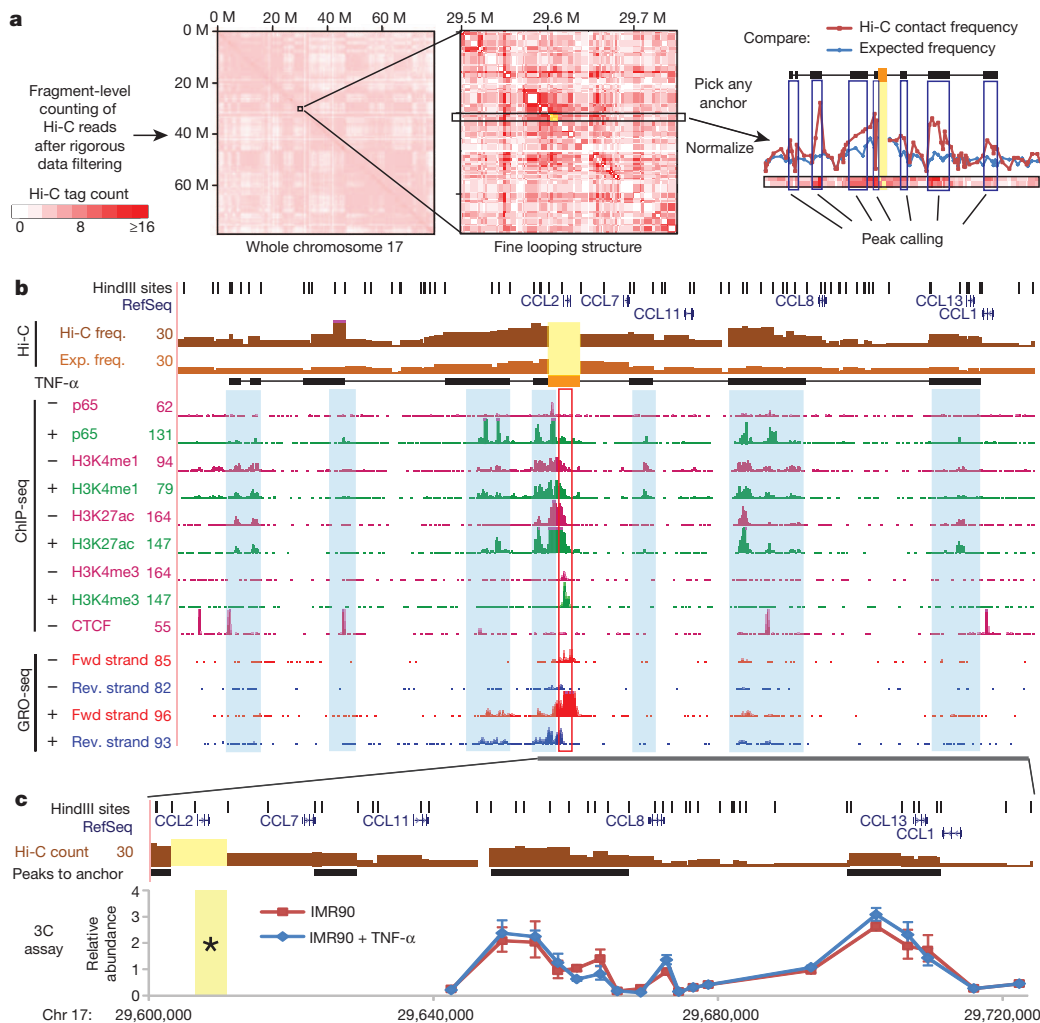
To accurately map at high resolution the chromatin interactions genome-wide, we devised an algorithm (Supplementary Fig. 12) to identify statistically significant looping interactions centred on a given genomic region from Hi-C contact matrix (Fig. 1a). We applied this method to the *CCL2* (chemokine (C-C motif) ligand 2) locus, and were able to determine the distal enhancers and CTCF (a chromatin organizer) binding sites interacting with the *CCL2* promoter (Fig. 1a, b). Our algorithm also identified a number of previously reported long-range chromatin interactions at the homeobox A (*HoxA*) gene cluster<sup>9</sup> and the sonic hedgehog (*SHH*) locus<sup>10</sup>, which were not readily observable from lower resolution analysis (Supplementary Figs 13 and 14). We then performed conventional 3C experiments to validate six pairs of long-range interactions identified at five different genes, and the results confirmed the reliability of our method (Fig. 1c and Supplementary Fig. 15).

We next applied the above algorithm to the 518,032 anchor regions in the human genome, with each containing one or a few HindIII restriction fragments (fragments shorter than 2 kilobases (kb) are merged) (Fig. 2a), and uncovered a total of 1,116,312 chromatin interactions with a false discovery rate (FDR) of 0.1 (Supplementary Data). We found that strong interactions supported by lower *P* values and higher contact frequencies are more reproducible between biological replicates (Supplementary Fig. 16). As interactions between loci separated by more than 2 megabases (Mb) are very rare (Fig. 2c), we limit our search to this genomic span. The sizes of the identified interacting DNA loci range from several hundred base pairs to over 50 kb, with a median of 10.5 kb (Fig. 2b). We were able to identify chromatin interactions that span a genomic distance from several hundred base pairs to over 1 million base pairs (Fig. 2c). Consistent with previous reports that the genome is partitioned into megabase-sized topological domains<sup>11–13</sup>, we found that a majority of the identified chromatin interactions in the IMR90 cells are located within the same topological domains (Fig. 2d and Supplementary Fig. 18).

We next characterized the chromatin interactions centred on the *cis*-elements annotated in the IMR90 cell genome<sup>14</sup> (Supplementary Data). Chromatin looping interactions are significantly enriched at *cis*-regulatory elements, particularly active promoters, enhancers and CTCF binding sites, and are rare at inactive transcription start sites (TSSs) or regions with repressive chromatin domains marked by H3K27me3 (Fig. 2e, f)<sup>15</sup>; notably, both active and poised enhancers (distinguished by the status of H3K27ac)<sup>16–18</sup> are found equally likely to engage looping interactions (Fig. 2e, f), raising the possibility that DNA looping could take place after priming of enhancers by H3K4me1 but before further activation<sup>19</sup>. Interestingly, the chromatin interactions

<sup>1</sup>Ludwig Institute for Cancer Research, 9500 Gilman Drive, La Jolla, California 92093, USA. <sup>2</sup>UCSD Medical Scientist Training Program, University of California, San Diego, School of Medicine, 9500 Gilman Drive, La Jolla, California 92093, USA. <sup>3</sup>UCSD Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA. <sup>4</sup>UCSD Biomedical Sciences Graduate Program, University of California, San Diego, School of Medicine, 9500 Gilman Drive, La Jolla, California 92093, USA. <sup>5</sup>Department of Cellular and Molecular Medicine, Institute of Genome Medicine, Moores Cancer Center, University of California, San Diego, School of Medicine, 9500 Gilman Drive, La Jolla, California 92093, USA.

\*These authors contributed equally to this work.



**Figure 1 | Fine mapping of chromatin interactions in IMR90 cells.** **a**, An illustration of the Hi-C data analysis procedure to identify regions interacting with a selected genomic region, such as the *CCL2* locus, highlighted in yellow (Supplementary Methods). **b**, Genome-browser shot of the *CCL2* locus showing the results from Hi-C, ChIP-seq and GRO-seq experiments. Each bar in the top two tracks are either Hi-C reads count (dark brown) or expected frequency (light brown) from a fragment to *CCL2* locus (highlighted in yellow and orange filled box). Black filled boxes are regions interacting with *CCL2* called by the peak calling algorithm same as the black filled boxes in **a**. Light blue shadows highlight the enhancer and CTCF locations from ChIP-seq data. Induction of *CCL2* by TNF- $\alpha$  is shown in the GRO-seq tracks. **c**, Validation of the DNA looping interactions with *CCL2* using 3C assays. Yellow, anchor fragments in Hi-C or 3C (asterisk). Error bar: s.d. from 3 technical replicates.

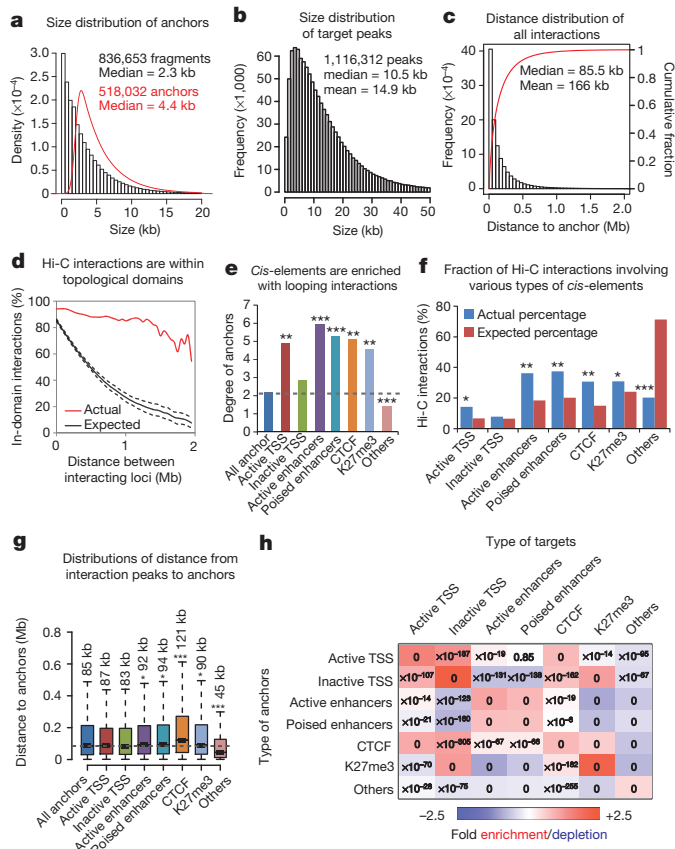
centred on CTCF binding sites tend to occur over a longer range than other types of *cis*-elements (Fig. 2g), confirming a recent result obtained from selected loci<sup>20</sup>. We also explored the spatial organization of *cis*-elements by examining preferential interaction between different classes of elements. Strongest enrichment was observed between H3K27me3 marked regions (Fig. 2h), consistent with the known compact three-dimensional structure at this type of repressive chromatin domains<sup>21</sup> (for example, Supplementary Fig. 14a). The inactive promoters tend to interact with regions depleted of enhancers but enriched for repressive mark H3K27me3 (Fig. 2h), whereas CTCF binding sites loop to both active and inactive promoters with no preference, as reported previously<sup>15</sup>. It is also interesting to observe that CTCF binding sites seem to prefer promoters over enhancers (Fig. 2h), suggesting a specific role for CTCF in organizing long-range chromatin interactions to promoters.

Looping interactions between *cis*-regulatory elements and gene promoters have been shown to be important for transcription regulation at a number of loci<sup>3</sup>. The genome-wide identification of chromatin interactions in the IMR90 cells allowed us to examine this concept systematically. We first focused on the looping interactions anchored to gene promoters, and denote the identified interacting sequences as promoter tethered regions (PTRs) (Fig. 3a). In IMR90 cells, we found 57,585 PTRs identifying 29,132 enhancer–promoter pairs involving 6,133 active promoters and 15,432 distal active enhancers (Supplementary Data). Only approximately 25% of enhancer–promoter pairs are within a 50-kb range, and approximately 57% span 100 kb or larger genomic distance, with a median distance of 124 kb (Fig. 3b). We

assigned 55% of distal enhancers to at least one active promoter, and 25% of enhancers to two or more active promoters (Fig. 3c, left panel). This result confirms previous observation that promoters and enhancers often form complex networks to regulate transcription<sup>15</sup>. We further reasoned that genes sharing common enhancers (denoted hub enhancers) are likely to have coordinated gene expression patterns. Indeed, genes sharing the same nuclear factor  $\kappa$ B (NF- $\kappa$ B)-responsive enhancers are more frequently induced together by TNF- $\alpha$  than expected by chance (Fig. 3d). As an example, *CKAP2L* and *IL1A* are induced simultaneously by TNF- $\alpha$  although lacking promoter-bound p65 peaks, and they share overlapping distal PTRs containing multiple NF- $\kappa$ B binding sites (Fig. 3e). Similar examples can be found in other gene clusters co-induced by TNF- $\alpha$  treatment (Supplementary Fig. 19). These results therefore provide a molecular mechanism for coordinated gene expression of neighbouring genes.

Interestingly, 46% of the active genes do not interact with any distal enhancer (Fig. 3c, right panel). Gene ontology analysis showed that these genes are enriched with housekeeping genes (Supplementary Fig. 20a). The remaining 54% of the active promoters demonstrate extensive looping interacting with enhancers (average 4.75 enhancers per gene, Fig. 3c, right panel), and they are enriched with genes related to biological pathways such as signal transduction (Supplementary Fig. 20b). This analysis suggests that housekeeping genes, despite being highly transcribed, do not engage a lot of distal regulatory elements. Conversely, genes involved in cell-specific functions are under extensive control of distal regulator sequences.

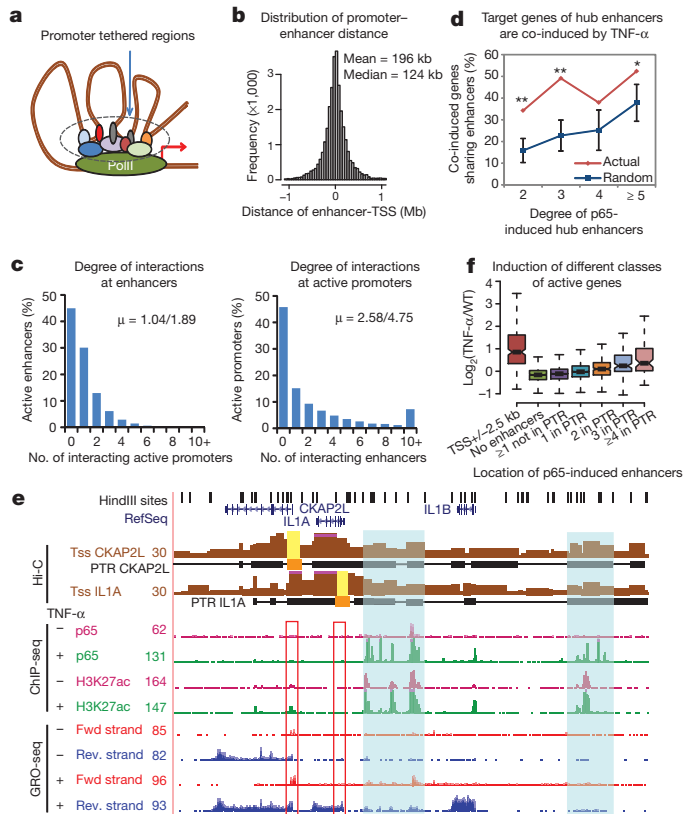




**Figure 2 | Characterization of the IMR90 chromatin interactome.**

**a**, Histogram showing the size distribution of all HindIII restriction fragments in human genome. The red curve plots the size distribution of all anchors used in this study. **b**, Size distribution of the genomic regions identified as interacting with all anchors. **c**, Distribution of genomic spans of all identified chromatin interactions. Histogram, frequency distribution; red curve, cumulative fraction with increasing distances. **d**, Fraction of chromatin looping interactions that fall within the same topological domains is plotted as a function of genomic distance between the two interacting loci (red curve). Black curve, expected fraction calculated from random shuffling locations of topological domains (100 iterations; dashed error lines, s.d.). **e**, Average number of peaks identified for anchors with different *cis*-elements. **f**, Percentages of all interactions involving various types of *cis*-elements (either anchor or target peak has the elements). In **e** and **f**, Z-scores were calculated comparing the actual values to simulation by randomly shuffling the locations of *cis*-elements (100 iterations, two-side Z-test). \* $Z > 50$ , \*\* $Z > 100$ , \*\*\* $Z > 150$ . **g**, Box plot showing the distance distribution from different types of anchors to their targets. Median distances are also labelled. *t*-statistics are computed comparing log-transformed distance between each type of anchors to all anchors as control (dash horizontal line). \* $t > 20$ , \*\* $t > 40$ , \*\*\* $t > 80$  (two-side *t*-test). **h**, Preferential interactions between different types of *cis*-elements. Heatmap shows the fold enrichment of different type of pair-wise combinations. *P* values are computed using hypergeometric test and denoted in each cell.

We next examined long-range looping interactions at transcriptional enhancers, focusing on those bound by the p65 subunit of NF- $\kappa$ B transcription factor. Using chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq), we identified 15,621 p65 binding sites in the genome after TNF- $\alpha$  treatment, 2,315 (14.8%) of which can be classified as 'active p65 binding sites' because they exhibit increased H3K27ac levels and enhancer RNA expression upon TNF- $\alpha$  signalling (Supplementary Figs 21 and 22). Consistent with their putative role in mediating transcriptional induction, these active p65 binding sites are enriched near TNF- $\alpha$ -dependent genes (Supplementary Fig. 22c). We next tested whether the long-range interactions between these p65 binding sites and their target promoters are correlated with transcriptional induction. Indeed, at the promoters that exhibit interactions



**Figure 3 | Identification and characterization of promoter-enhancer interactions in IMR90 cells.** **a**, A schematic of promoter tethered regions (PTRs). **b**, Distribution of distances between the promoters and enhancers found within the PTRs. **c**, Bar charts show the degrees of interactions. Left, percentage of enhancers that are looped to active promoters found in PTRs. The first value of  $\mu$  is the mean degree of promoter interactions for all enhancers; the second value is the mean degree of promoter interactions for enhancers interacting with at least one promoter. Right, distribution of degrees of enhancer interactions for active promoters. **d**, Genes sharing enhancers are co-regulated. Among the target genes of TNF- $\alpha$ -responsive hub enhancers, proportion of co-induced gene pairs (greater than twofold) are plotted and compared to random simulation (100 iterations, two-side Z-test, \* $P < 0.05$ , \*\* $P < 0.01$ , error bar, s.d.). **e**, Genome browser snapshots showing the virtual 3C plots of the *CKAP2L* and *IL1A* promoters. **f**, Box plots comparing the induction of different gene groups based on the location of p65-induced enhancers. 'No enhancers', genes with no such enhancers within 2 Mb; '≥1 not in PTR', p65-induced enhancers are within 2 Mb but not in PTR.

with one or more active p65 binding sites, significantly higher levels of transcriptional induction were observed than the promoters that do not interact with distal p65 binding sites (Fig. 3f), suggesting that the identified long-range chromatin interactions may have a key role in transcriptional regulation of the TNF- $\alpha$ -inducible genes.

The high-resolution map of chromatin interactions may also improve the prediction of target genes of distal enhancers. Currently, a common practice is to assign distal enhancers to their nearest promoters, assuming one enhancer is linked to just one target gene (proximity approach). However, this approach cannot explain all of the 828 TNF- $\alpha$ -responsive genes. We found that 331 (40%) of these genes have one or more p65 binding sites within 2.5 kb of their promoters, and 362 genes of the remainder can be assigned to one or more NF- $\kappa$ B binding sites by proximity approach, leaving still 135 TNF- $\alpha$ -induced genes unexplained. Using a recently published enhancer-promoter connection map<sup>22</sup> based on correlated chromatin features across diverse tissues or cell types, we were able to link 10 of the 135 unexplained TNF- $\alpha$ -inducible genes to distal NF- $\kappa$ B binding sites. Using the chromatin interactome map, 74 (55%) of the unexplained genes can be assigned to a NF- $\kappa$ B binding site (Supplementary Fig. 23). This result illustrates that the chromatin

interactome map described here could be valuable for the study of long-range regulation of gene expression by transcription factors.

We found no obvious alterations of megabase topological domains<sup>12</sup> in IMR90 cells after TNF- $\alpha$  treatment (for example, Supplementary Fig. 13). As previous studies have shown that gene activation by enhancers is accompanied by alteration of chromatin interactions<sup>3,23,24</sup>, we expected that at shorter distance, binding of NF- $\kappa$ B to enhancers would induce looping interactions that bring the distal enhancers to proximity with target genes. To our surprise, we found that at the vast majority of TNF- $\alpha$  responsive enhancers, there is little change of DNA looping after treatment (Fig. 4a). These results suggest that in general, enhancer–promoter interactions already form in untreated cells; and these pre-existing DNA-structures are not significantly altered by transient activation or repression of enhancers. 3C assays confirm that DNA looping exists at several loci before TNF- $\alpha$  treatment (Fig. 1e and Supplementary Fig. 15a–c). We further compared the normalized reads count anchored to the TNF- $\alpha$ -activated enhancers at different ranges of genomic distances (Fig. 4b and Supplementary Methods). Consistent with the results in Fig. 4a, transient activation of p65-bound enhancers does not lead to significant changes in chromatin interactions (Fig. 4b). By contrast, the chromatin interactions (especially within a short genomic distance) at cell-type-specific enhancers are highly variable between cell types (Fig. 4c, d), suggesting the existence of specific chromatin interaction structures at cell-type-specific enhancers. Interestingly, the discrepancy between signal-dependent and cell-type-specific enhancers is well correlated with the levels of H3K4me1 at those dynamically regulated enhancers: despite the quick induction of the H3K27ac mark at TNF- $\alpha$ -responsive enhancers, strength of H3K4me1 signal is largely unchanged (Fig. 4e), but on the other hand, cell-type-specific enhancers have highly cell-specific H3K4me1 occupancy (Fig. 4f).

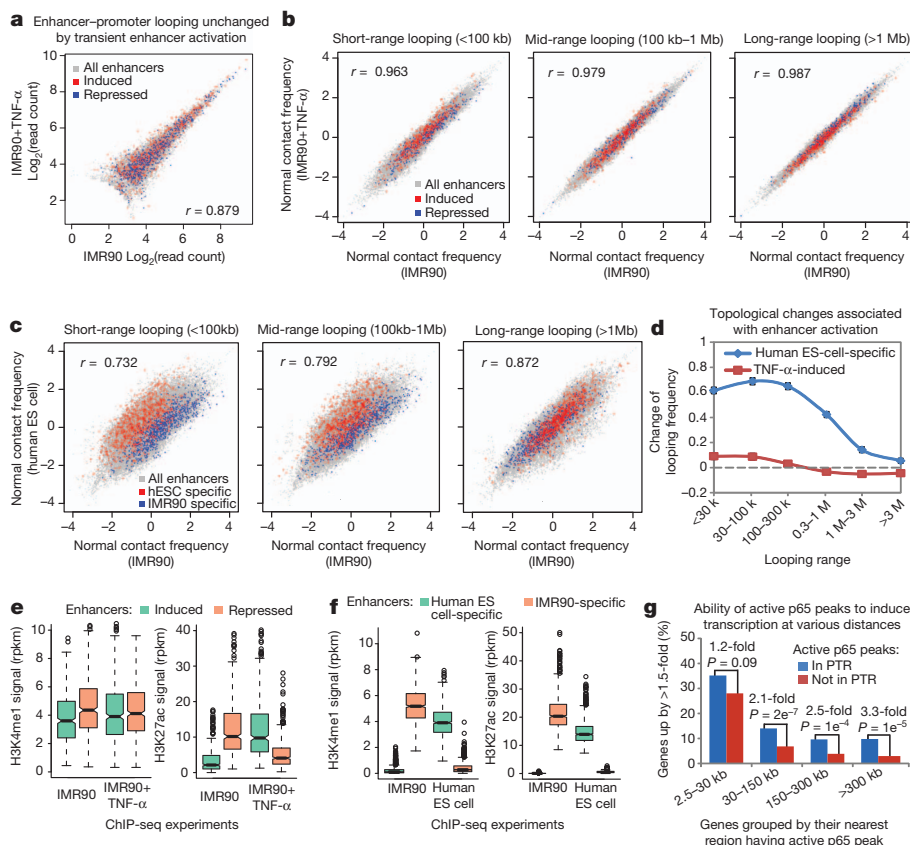
Recently, pre-existing promoter–enhancer looping was observed at several loci induced by p53, FOXO3 and glucocorticoid receptor using the 4C (chromosome conformation capture combined with sequencing)

approach<sup>20,25,26</sup>. Our genome-wide analysis of chromatin interactome maps in IMR90 cells suggests that this is likely to be a common rule rather than a special case. To further demonstrate its generality, we examined six additional promoter–enhancer pairs by 3C assays in four different cell types (IMR90, HUVEC, MCF7 and LNCaP cells) under different stimuli (IFN- $\gamma$ , TNF- $\alpha$ ,  $\beta$ -oestradiol and 5 $\alpha$ -dihydrotestosterone, respectively). In all of these examples, we found evident pre-existing promoter–enhancer contacts, and the looping interactions are largely unchanged after enhancer activation and target gene induction (Supplementary Figs 15d and 24). Our results predict that pre-existing chromatin looping interactions could dictate the spectrum of the target genes for a transcription factor even before it is activated. Indeed, p65 binding sites looping to the promoters before induction are much more likely to result in transcriptional activation of the linked gene than otherwise (Fig. 4g). This trend is particularly obvious when the p65-binding sites are located far from the linked promoters. On the basis of this observation, we conclude that pre-existing DNA-looping interactions between enhancers and promoters allow a ubiquitous, signal-dependent transcription factor to affect a selected set of genes in a cell-type-specific manner.

In summary, we have demonstrated that enhancer–promoter interactions already form in each cell type before the binding of signal dependent transcription factors, and they undergo relatively few changes during transient transcriptional activation. Several recent genome-wide studies have revealed that in different cell types, the repertoire of specific enhancers provides a unique context for the activation of different transcriptional programs in response to signal-dependent transcription factors including NF- $\kappa$ B<sup>19,27–30</sup>. Here, our results further suggest that targets of cell-specific enhancers are already hardwired into the chromatin architecture in each cell lineage. We therefore propose that cell-type-specific looping structure, by controlling the accessibility of the enhancers to their specific targets, may form an additional layer of regulation in determining the distinct transcription programs in different cell types.

**Figure 4 | The higher order chromatin structure in IMR90 cells is stable during transient TNF- $\alpha$  signalling.**

**a**, A scatter plot comparing reads count at PTRs before and after TNF- $\alpha$  treatment. Grey dots are data for PTRs involving all enhancers, red or blue dots are PTRs involving the top 500 induced or repressed enhancers, respectively.  $r$ , Pearson's correlation calculated from all data points. **b**, Scatter plots compare normalized contact frequencies (Supplementary Methods) of all enhancers before and after 1 h TNF- $\alpha$  stimulation. **c**, Scatter plots compare contact frequencies of enhancers in IMR90 and hESC cells. Coloured points represent the top 2,000 human embryonic stem cell (ES-cell)-specific (red) or IMR90-specific (blue) enhancers defined by H3K27ac mark. **d**, The relative change of contact frequency (comparing to untreated IMR90 cells) at human ES-cell-specific or TNF- $\alpha$ -induced enhancers are plotted in line graphs. **e**, Box plots of H3K4me1 and H3K27ac signals on enhancers that show increased or decreased H3K27ac signals after TNF- $\alpha$  treatment. rpkms, reads per kilobase per million. **f**, Boxplots of H3K4me1 and H3K27ac ChIP-seq signals at the IMR90- or hESC-specific enhancers. **g**, Bar charts showing that NF- $\kappa$ B binding sites within PTRs are more likely to activate target genes than those outside PTRs. In this figure, PTRs are identified using the Hi-C data from untreated IMR90 cells. The  $P$  values are calculated using hypergeometric test.



## METHODS SUMMARY

Hi-C experiments were performed in human primary IMR90 fibroblasts. ChIP-seq, global run-on sequencing (GRO-seq), or RNA-seq libraries were also generated from IMR90, HUVEC, MCF7 or LNCaP cells and sequenced on the Illumina Hi-Seq2000 platform. All the reads were mapped to reference human genome (hg18). More information about the experiments and detailed descriptions of Hi-C data analysis pipeline, including data filtering, normalization, statistical modelling and interaction calling can be found in Supplementary Methods.

Received 30 December 2012; accepted 6 September 2013.

Published online 20 October 2013.

1. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
2. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
3. Smallwood, A. & Ren, B. Genome organization and long-range regulation of gene expression by enhancers. *Curr. Opin. Cell Biol.* **25**, 387–394 (2013).
4. Dekker, J., Marti-Renom, M. A. & Mirny, L. A. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nature Rev. Genet.* **14**, 390–403 (2013).
5. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
6. Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nature Methods* **9**, 999–1003 (2012).
7. de Wit, E. & de Laat, W. A decade of 3C technologies: insights into nuclear organization. *Genes Dev.* **26**, 11–24 (2012).
8. Mercer, T. R. *et al.* DNase I-hypersensitive exons colocalize with promoters and distal regulatory elements. *Nature Genet.* **45**, 852–859 (2013).
9. Noordermeer, D. *et al.* The dynamic architecture of *Hox* gene clusters. *Science* **334**, 222–225 (2011).
10. Lettice, L. A. *et al.* Disruption of a long-range *cis*-acting regulator for *Shh* causes preaxial polydactyly. *Proc. Natl Acad. Sci. USA* **99**, 7548–7553 (2002).
11. Zhang, Y. *et al.* Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* **148**, 908–921 (2012).
12. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
13. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
14. Hawkins, R. D. *et al.* Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* **6**, 479–491 (2010).
15. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109–113 (2012).
16. Creighton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl Acad. Sci. USA* **107**, 21931–21936 (2010).
17. Hawkins, R. D. *et al.* Dynamic chromatin states in human ES cells reveal potential regulatory sequences and genes involved in pluripotency. *Cell Res.* **21**, 1393–1409 (2011).
18. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283 (2011).
19. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
20. Phillips-Cremins, J. E. *et al.* Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013).
21. Francis, N. J., Kingston, R. E. & Woodcock, C. L. Chromatin compaction by a polycomb group protein complex. *Science* **306**, 1574–1577 (2004).
22. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
23. Ong, C. T. & Corces, V. G. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nature Rev. Genet.* **12**, 283–293 (2011).
24. Schoenfelder, S., Clay, I. & Fraser, P. The transcriptional interactome: gene expression in 3D. *Curr. Opin. Genet. Dev.* **20**, 127–133 (2010).
25. Melo, C. A. *et al.* eRNAs are required for p53-dependent enhancer activity and gene transcription. *Mol. Cell* **49**, 524–535 (2013).
26. Tan, P. Y. *et al.* Integration of regulatory networks by NKX3-1 promotes androgen-dependent prostate cancer survival. *Mol. Cell Biol.* **32**, 399–414 (2012).
27. Jin, F., Li, Y., Ren, B. & Natarajan, R. P. U. 1 and C/EBP(alpha) synergistically program distinct response to NF- $\kappa$ B activation through establishing monocyte specific enhancers. *Proc. Natl Acad. Sci. USA* **108**, 5290–5295 (2011).
28. John, S. *et al.* Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nature Genet.* **43**, 264–268 (2011).
29. Mullen, A. C. *et al.* Master transcription factors determine cell-type-specific responses to TGF-beta signaling. *Cell* **147**, 565–576 (2011).
30. Jin, F., Li, Y., Ren, B. & Natarajan, R. Enhancers: multi-dimensional signal integrators. *Transcription* **2**, 226–230 (2011).

Supplementary Information is available in the online version of the paper.

**Acknowledgements** We thank C. K. Glass for sharing the GRO-seq protocol, and S. Kuan and L. Edsall for assistance with high-throughput DNA sequencing and the initial processing. This work is supported by funds from the Ludwig Institute for Cancer Research, the California Institute of Regenerative Medicine (RN2-00905) and US National Institutes of Health (P50 GM085764-03 and U01 ES017166).

**Author Contributions** Y.L., F.J. and B.R. designed the studies. Y.L. conducted most of the experiments; F.J. carried out the data analysis; J.R.D., Z.Y., A.Y.L., C.Y., A.D.S. and C.E. contributed to the experiments; S.S. contributed to the data analysis; F.J., Y.L. and B.R. prepared the manuscript.

**Author Information** All sequencing data described in this study have been deposited to GEO under the accession number GSE43070. Some sequencing data used in this study were previously published and accession numbers can be found in Supplementary Methods. All chromatin interactions called in IMR90 cells can be found in Supplementary Data. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.R. ([biren@ucsd.edu](mailto:biren@ucsd.edu)).



# Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs

Ron O. Dror<sup>1\*</sup>, Hillary F. Green<sup>1\*</sup>, Celine Valant<sup>2\*</sup>, David W. Borhani<sup>1\*</sup>, James R. Valcourt<sup>1</sup>, Albert C. Pan<sup>1</sup>, Daniel H. Arlow<sup>1</sup>, Meritxell Canals<sup>2</sup>, J. Robert Lane<sup>2</sup>, Raphaël Rahmani<sup>3</sup>, Jonathan B. Baell<sup>3</sup>, Patrick M. Sexton<sup>2</sup>, Arthur Christopoulos<sup>2</sup> & David E. Shaw<sup>1,4</sup>

**The design of G-protein-coupled receptor (GPCR) allosteric modulators, an active area of modern pharmaceutical research, has proved challenging because neither the binding modes nor the molecular mechanisms of such drugs are known<sup>1,2</sup>. Here we determine binding sites, bound conformations and specific drug-receptor interactions for several allosteric modulators of the M2 muscarinic acetylcholine receptor (M2 receptor), a prototypical family A GPCR, using atomic-level simulations in which the modulators spontaneously associate with the receptor. Despite substantial structural diversity, all modulators form cation- $\pi$  interactions with clusters of aromatic residues in the receptor extracellular vestibule, approximately 15 Å from the classical, 'orthosteric' ligand-binding site. We validate the observed modulator binding modes through radioligand binding experiments on receptor mutants designed, on the basis of our simulations, either to increase or to decrease modulator affinity. Simulations also revealed mechanisms that contribute to positive and negative allosteric modulation of classical ligand binding, including coupled conformational changes of the two binding sites and electrostatic interactions between ligands in these sites. These observations enabled the design of chemical modifications that substantially alter a modulator's allosteric effects. Our findings thus provide a structural basis for the rational design of allosteric modulators targeting muscarinic and possibly other GPCRs.**

A third of marketed drugs act on GPCRs, generally by binding at the orthosteric site in competition with the endogenous ligands that naturally regulate receptor signalling<sup>3</sup>. There is considerable interest, however, in drugs that bind at distinct, allosteric sites. Such allosteric modulators may achieve selectivity between receptor subtypes that possess nearly identical orthosteric sites<sup>1,4</sup>. In addition, because they act primarily by altering the affinity or efficacy of endogenous ligands, allosteric modulators may provide a means to fine-tune cellular responses to the body's natural signalling patterns<sup>4</sup>.

Despite recent advances in GPCR crystallography<sup>5–7</sup>, no structure has been determined of a GPCR bound to a drug-like allosteric modulator. Moreover, the molecular mechanisms by which such modulators affect GPCR signalling may depend on dynamical properties that would not be evident from a single static structure.

To address these structural and mechanistic questions, we performed atomic-level molecular dynamics simulations of several allosteric modulators binding spontaneously to the M2 receptor, a long-serving model system for allosteric modulation of GPCRs<sup>8</sup>. Such simulations have previously proved capable of capturing the complete process of drug binding, with a ligand exploring a receptor's surface and 'discovering' the binding conformation at crystallographic accuracy, without previous knowledge of the binding site<sup>9–11</sup>. Our simulations, which included lipids, water and ions surrounding the receptor, allowed

conformational changes in both the receptor and the modulators, and made no assumptions about where the modulators might bind.

Figure 1a depicts the simulated binding process of a prototypical allosteric modulator, C<sub>7</sub>/3-phth, to an unliganded human M2 receptor (cf. Supplementary Fig. 1 and Supplementary Tables 1 and 2). C<sub>7</sub>/3-phth settled into a stable and reproducible bound pose in the extracellular vestibule, a region approximately 15 Å from the orthosteric site (Fig. 1b, Supplementary Video 1, Supplementary Fig. 2 and Supplementary Table 3). The C<sub>7</sub>/3-phth core—the two ammonium groups and the heptyl linker—remained tightly bound to the receptor, whereas the distal phthalimide groups were highly mobile (Supplementary Fig. 3). The close structural analogue dimethyl-W84 bound in a very similar fashion (Fig. 1c). The binding modes observed for C<sub>7</sub>/3-phth and dimethyl-W84 are consistent with extensive mutagenesis data: all mutated residues that reduce the binding affinity of C<sub>7</sub>/3-phth, dimethyl-W84 or related bis-ammonium alkane ligands by more than fivefold<sup>12–14</sup> formed stable contacts with the relatively immobile C<sub>7</sub>/3-phth and dimethyl-W84 cores (Fig. 1b and Supplementary Fig. 4).

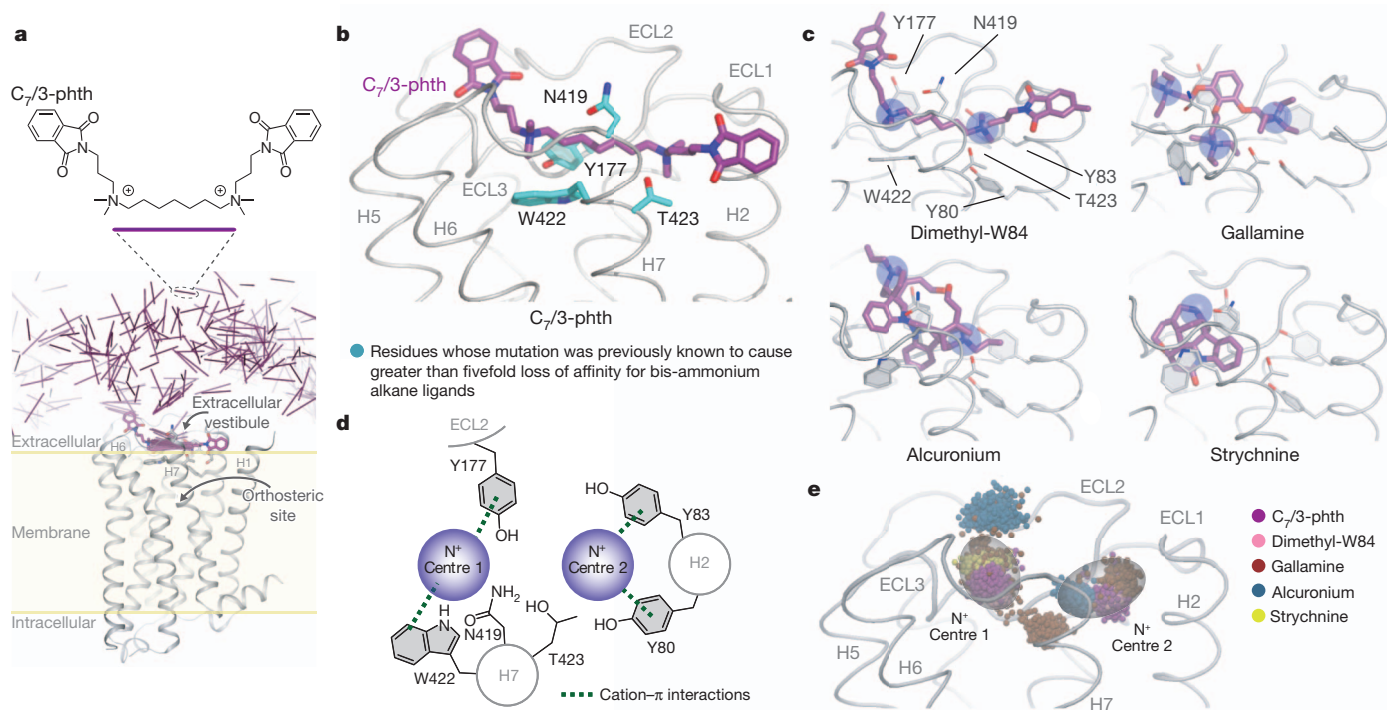
We extended our binding simulations to include structurally diverse and clinically relevant drugs, including the neuromuscular-blocking drugs gallamine and alcuronium and the convulsant strychnine. These modulators are known to compete with one another, C<sub>7</sub>/3-phth and dimethyl-W84 for binding to the M2 receptor, and are thus believed to bind to the same region of the receptor<sup>15</sup>.

In our simulations, all these allosteric modulators not only bound to the same site—the extracellular vestibule—but also made similar binding interactions, despite their structural diversity (Fig. 1c and Supplementary Fig. 5). Ammonium groups from each ligand occupied one or both of two common 'binding centres' in the extracellular vestibule. These centres are each defined by a pair of aromatic residues—an 'aromatic box'—that form cation- $\pi$  interactions with the ammonium group (Fig. 1d, e): Tyr 177<sup>ECL2</sup> and Trp 422<sup>7,35</sup> at centre 1, and Tyr 80<sup>2,61</sup> and Tyr 83<sup>2,64</sup> at centre 2. (Superscripts refer to Ballesteros-Weinstein residue numbering<sup>16</sup>, with ECL2 denoting the second extracellular loop.)

The specific poses and receptor-drug interactions we observed, although completely in accord with extensive previous mutagenesis data (Supplementary Fig. 4) and the widely held belief that these allosteric modulators bind in the vicinity of the extracellular vestibule<sup>2,17,18</sup>, generally differ from earlier predictions. Previous studies have suggested, for example, that the abundant aromatic residues in the vestibule interact with the aromatic groups common to many allosteric modulators<sup>12,14</sup>. We found that these residues instead formed cation- $\pi$  interactions with modulator ammonium groups; modulator aromatic groups, when present, formed non-specific hydrophobic contacts with the receptor. Indeed, hexamethonium, devoid of aromatic groups, bound in a pose analogous to that of the C<sub>7</sub>/3-phth core in simulation

<sup>1</sup>D. E. Shaw Research, 120 West 45th Street, 39th Floor, New York, New York 10036, USA. <sup>2</sup>Drug Discovery Biology, Monash Institute of Pharmaceutical Sciences and Department of Pharmacology, Monash University, 399 Royal Parade, Parkville, 3052 Victoria, Australia. <sup>3</sup>Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, 399 Royal Parade, Parkville, 3052 Victoria, Australia. <sup>4</sup>Center for Computational Biology and Bioinformatics, Columbia University, New York, New York 10032, USA.

\*These authors contributed equally to this work.



**Figure 1 | Structurally diverse allosteric modulators bind spontaneously to the human M2 muscarinic acetylcholine receptor in simulation, revealing a common binding mode.** **a**, *C*<sub>7</sub>/3-phth diffuses freely before binding stably in the M2 receptor extracellular vestibule. The *C*<sub>7</sub>/3-phth position at various times is represented by a stick connecting its two ammonium groups. (Top: *C*<sub>7</sub>/3-phth structure.) **b**, Typical bound pose of *C*<sub>7</sub>/3-phth (purple). Residues known to reduce binding greater than fivefold upon mutation (cyan) all contact *C*<sub>7</sub>/3-phth more than 90% of the time after it binds. **c**, Typical bound poses for dimethyl-W84, gallamine, strychnine and alcuronium. Ammonium groups are

highlighted (blue disks). **d**, Schematic representation of the ammonium binding centres. **e**, Bound locations of modulator ammonium groups (spheres). *C*<sub>7</sub>/3-phth, dimethyl-W84 and gallamine each occupy centres 1 and 2 (grey ellipsoids); strychnine occupies only centre 1, and alcuronium only centre 2. The position of each centre varies slightly depending on the bound modulator because the surrounding residues reposition to accommodate the modulator. The third gallamine ammonium and the second of alcuronium lie outside the two centres (alcuronium cannot occupy both simultaneously owing to its geometry).

(Supplementary Fig. 6). We also found that modulator cationic ammonium groups did not form stable contacts with anionic receptor residues such as those in the widely studied ECL2 ‘EDGE’ motif<sup>3</sup> (Supplementary Fig. 7), in contrast to such direct recognition of the ammonium group of orthosteric ligands by the conserved anionic residue Asp 103<sup>3,32</sup> (ref. 5).

To validate these computationally determined binding modes, we experimentally measured the binding of *C*<sub>7</sub>/3-phth and gallamine to several new M2 receptor mutants designed, on the basis of simulations, either to increase or to decrease modulator affinity (Fig. 2 and Supplementary Fig. 8). To increase affinity, we introduced aromatic residues near centres 1 and 2, to strengthen the cation– $\pi$  interactions (Fig. 2a), or anionic residues near centre 2, to form favourable ionic interactions (Fig. 2b). To decrease affinity, we introduced cationic residues near either centre (Fig. 2c). In all cases, affinity increased or decreased as predicted (Fig. 2d and Supplementary Figs 9–12). Additional mutagenesis results suggest that centre 1 contributes more strongly than centre 2 to *C*<sub>7</sub>/3-phth and gallamine binding, in accord with computational findings (Supplementary Figs 13 and 14).

Next, we examined the mechanism of allosteric modulation, in particular, the ability of allosteric modulators to increase or decrease the affinity of orthosteric ligands. Because the M2 receptor crystal structure represents an inactive, antagonist-bound state, we focused on modulation of antagonist binding.

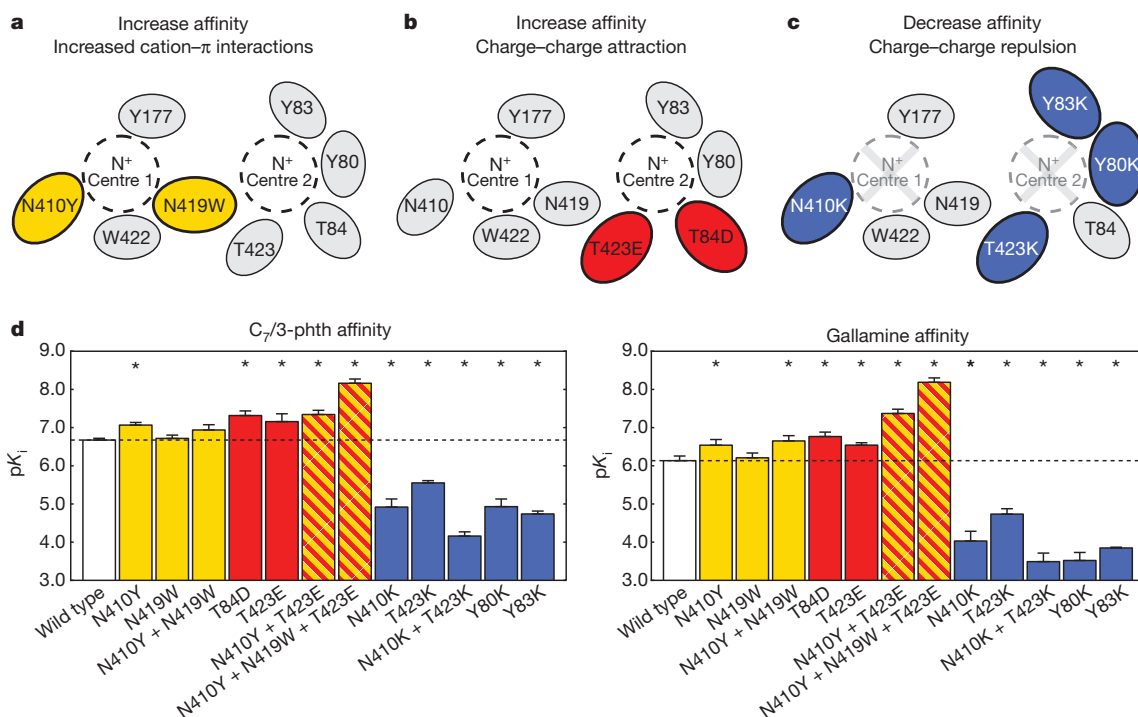
*C*<sub>7</sub>/3-phth, dimethyl-W84 and gallamine are negative allosteric modulators (NAMs) of the archetypal orthosteric antagonist *N*-methylscopolamine (NMS): their binding weakens NMS affinity, and vice versa. Alcuronium and strychnine, by contrast, are positive allosteric modulators (PAMs) of NMS: their binding increases NMS affinity. Simulations of modulators binding to unliganded and NMS-bound receptors were consistent

with these experimental observations (Fig. 3a, Supplementary Figs 2, 15 and 16 and Supplementary Table 3). Although *C*<sub>7</sub>/3-phth bound consistently in a stable, reproducible pose in simulations that did not include an orthosteric ligand, for example, it bound only transiently to an NMS-occupied receptor (Fig. 3a).

Our simulations revealed two mechanisms that seem to contribute to the cooperativity of orthosteric and allosteric ligand binding. First, electrostatic repulsion between a cationic orthosteric ligand (for example, NMS) and a cationic allosteric modulator weakens the binding of one in the presence of the other. The degree of repulsion depends on the total charges of both the modulator and the orthosteric ligand, and on the locations of those charges when the ligands are bound (Fig. 3b). Less cationic NAMs, or those whose positive charges are further from NMS, experience less repulsion. Electrostatic repulsion seems to account for most of the experimentally observed negative cooperativity of the NAMs we studied (Supplementary Table 4). Indeed, replacement of a cationic nitrogen atom with a neutral silicon atom was found to shift the cooperativity of *C*<sub>7</sub>/3-phth in the positive direction<sup>19</sup>.

Electrostatic repulsion, however, cannot explain the positive cooperativity of cationic PAMs. Indeed, on the basis of electrostatics alone, the NAM *C*<sub>7</sub>/3-phth and the PAM alcuronium would be predicted to have identical, not opposite, cooperativities.

Our simulations suggest a second operative mechanism underlying cooperativity: the conformations of the orthosteric and allosteric sites are coupled such that the presence of a ligand in one affects the shape of the other (Fig. 3c and Supplementary Fig. 17). In simulations with NMS bound in the orthosteric site, both the allosteric and orthosteric sites maintained the wide-open conformations observed crystallographically in the presence of 3-quinuclidinyl benzilate (QNB), another orthosteric antagonist. By contrast, in simulations with no ligand present,



**Figure 2 | Experimental validation of computationally derived binding modes.** **a–c**, Mutagenesis predictions; observed binding poses, and simulations of mutants, indicate that certain mutations should increase or decrease  $C_7/3$ -phth and gallamine binding affinity. **a**, Aromatic residues (N410Y, N419W) contribute extra cation- $\pi$  interactions at centre 1, increasing affinity. **b**, Anionic residues (T423E, T84D) attract an ammonium group at centre 2, increasing affinity. **c**, Cationic residues (N410K, N423K, Y80K, Y83K) near centres 1 and 2

repel or compete with cationic modulators, decreasing affinity. **d**, Mutagenesis results; radioligand binding experiments ( $[^3H]$ NMS displacement), performed on whole cells expressing wild-type and mutant human M2 receptors, confirmed computational predictions. Data represent mean  $\pm$  s.e.m. of three to five independent experiments (see Supplementary Table 6). Asterisks indicate a significant difference from the wild type ( $P < 0.05$ ).

both the orthosteric and allosteric sites assumed a narrower conformation, with ECL2 shifting approximately 2.5 Å towards ECL3 as helix 5 shifted slightly into the region vacated by the orthosteric ligand (Fig. 3c).

Interestingly, binding of the wide, bulky PAM alcuronium tended to force the allosteric and orthosteric sites into their open conformations, even when the orthosteric site was empty (Supplementary Figs 17 and 18). This conformational change seems to contribute to the positive cooperativity between NMS and alcuronium: binding of either one shifts the receptor into the conformation with higher affinity for the other. By contrast, the narrow NAM  $C_7/3$ -phth can bind to either the wide or narrow vestibule conformation; in fact, it forms more extensive contacts when bound to the narrow conformation, which is favoured when the orthosteric site is empty. Strychnine's weak positive cooperativity seems to result from a mild preference for a more open vestibule conformation coupled with the weakest electrostatic repulsion of the modulators studied. Although our data do not rule out other possible mechanisms, the combination of these two effects—sterics and electrostatics—seems sufficient to explain positive and negative modulation by these allosteric modulators.

On the basis of this mechanistic understanding, we predicted that one could shift the cooperativity of a modulator in the positive direction by modifying the modulator such that it could bind only to a wider conformation of the extracellular vestibule. To test this prediction, we designed and synthesized a bulky  $C_7/3$ -phth variant, 4P- $C_7/3$ -phth, that bound in the same pose as  $C_7/3$ -phth in test simulations but required a wider vestibule to do so (Fig. 3d and Supplementary Fig. 19). Indeed, radioligand binding experiments showed that, in comparison with  $C_7/3$ -phth, 4P- $C_7/3$ -phth has significantly reduced negative—that is, approximately fourfold more positive—cooperativity with NMS (Fig. 3d). 4P- $C_7/3$ -phth also has approximately 50-fold higher binding affinity than  $C_7/3$ -phth, suggesting that this modification did not disrupt the binding mode (Supplementary Fig. 20).

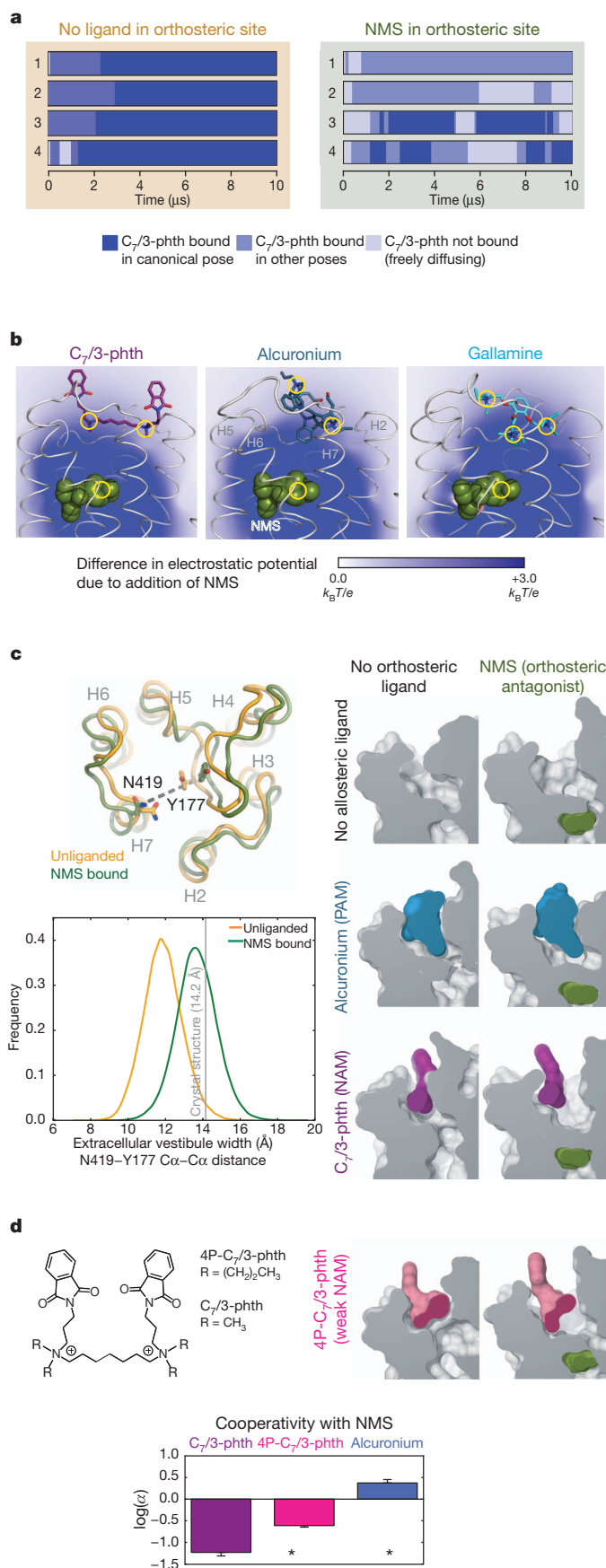
Allosteric modulation of agonist binding may rely on similar mechanisms. As muscarinic agonists, including the endogenous agonist acetylcholine, are generally cations, electrostatic interactions between modulators and agonists should mimic those seen with antagonists. The smaller size typical of agonists, and the orthosteric site contraction seen on activation of other family A GPCRs<sup>7,20</sup>, suggests that PAMs for agonists may tend to be smaller than PAMs for antagonists.

Effective structure-based design of allosteric modulators must contend with the flexibility of the extracellular vestibule<sup>21</sup>, which exhibits various conformations preferred by different modulators. In particular, computational screening approaches based on a single rigid receptor structure may identify only a subset of potent allosteric modulators<sup>22,23</sup>. Indeed, our attempts to predict modulator binding poses using standard docking protocols with the M2 receptor crystal structure yielded results substantially different from those observed in molecular dynamics simulations and in worse agreement with experimental data (Supplementary Fig. 21).

Our results provide a structural basis for the subtype selectivity of muscarinic allosteric modulators<sup>1</sup>. The typically lower affinity of modulators for the M3 receptor<sup>4</sup>, for example, is probably caused by the unique presence of a cationic residue, Lys 523<sup>7,32</sup>, within 8 Å of centres 1 and 2 (Supplementary Fig. 22). Several other poorly conserved residues that form contacts with the bound allosteric modulators might also be exploited to achieve a desired selectivity profile (Supplementary Table 5).

All the allosteric modulators we studied—PAMs and NAMs—typically retard both binding and dissociation of orthosteric ligands<sup>24</sup>. Confirming a previous hypothesis<sup>9,17</sup>, our simulations indicate that orthosteric ligands cannot easily gain access to, or egress from, their binding site when a modulator is bound in the extracellular vestibule, because the vestibule also typically serves as a way station along the binding pathway of orthosteric ligands. This explains how NAMs





**Figure 3 | Mechanisms of positive and negative allosteric modulation.**

**a**, Allosteric modulator binding affinity depends on orthosteric site occupancy. Each bar represents a simulation of C<sub>7</sub>/3-phth binding to unliganded (left) or NMS-bound (right) M2 receptor. C<sub>7</sub>/3-phth bound stably in the canonical pose to the unliganded receptor, but frequently shifted pose or dissociated entirely when NMS was bound. **b**, Electrostatic interactions between ligands contribute to cooperativity. Blue shading denotes the change in electrostatic potential due to addition of NMS. Modulator ammonium groups are circled (yellow). Electrostatic repulsion with NMS is similar for C<sub>7</sub>/3-phth and alcuronium, but substantially larger for gallamine, which positions its third ammonium group particularly close to the orthosteric site (see Supplementary Table 4). **c**, Conformational coupling between sites contributes to cooperativity. Left: with no allosteric modulator bound, the extracellular vestibule of NMS-bound receptor is wider than that of unliganded receptor (molecular snapshots and histograms from 10-μs simulations). Right: cross-sectional views from simulations without (left) or with (right) NMS in the orthosteric site, and with no modulator (top), alcuronium (middle) or C<sub>7</sub>/3-phth (bottom) in the allosteric site. Alcuronium binding requires a wide vestibule, whereas C<sub>7</sub>/3-phth can bind to a wide or narrow vestibule. **d**, We designed 4P-C<sub>7</sub>/3-phth (left) computationally to adopt the C<sub>7</sub>/3-phth binding pose but require a wider vestibule (right), rendering its cooperativity more positive (less negative). [<sup>3</sup>H]NMS binding experiments (bottom) indicate that the cooperativity of 4P-C<sub>7</sub>/3-phth is approximately fourfold less negative than that of C<sub>7</sub>/3-phth. Data represent mean ± s.e.m. of four independent experiments (see Supplementary Table 7). Asterisks indicate a significant difference from C<sub>7</sub>/3-phth ( $P < 0.05$ ).

may slow dissociation of an orthosteric ligand despite weakening its equilibrium affinity.

Although numerous allosteric modulators seem to bind in or near the extracellular vestibules of muscarinic receptors and other GPCRs<sup>25,26</sup>, certain modulators seem to bind elsewhere<sup>27,28</sup>. Further computational and experimental investigation is necessary to determine the extent to which the molecular basis of allosteric modulation is conserved across GPCRs. The techniques we have demonstrated here might also be used to identify allosteric mechanisms operative in other protein families.

## METHODS SUMMARY

Molecular dynamics simulations of the human M2 receptor, embedded in a hydrated palmitoylphosphatidylcholine bilayer with all atoms represented explicitly, were performed on Anton<sup>29</sup> using the CHARMM force field<sup>30</sup>. Simulations began with the receptor in the crystallographic conformation (Protein Data Bank accession number 3UON; ref. 5), but in most cases the co-crystallized antagonist (QNB) was either removed or replaced by NMS (Supplementary Fig. 23). In binding simulations, a single ligand was placed in the aqueous region away from the receptor. No artificial forces were applied to the ligand, which generally diffused extensively about the simulated volume before binding.

Mutant receptors were constructed using standard cloning techniques and expressed as Chinese hamster ovary (CHO) stable cell lines. Saturation binding experiments were performed on whole cells using [<sup>3</sup>H]NMS (0.003–3 nM), in the absence or presence of atropine (100 μM) at 22 °C. Inhibition binding and dissociation kinetics experiments, also on whole cells, were performed using 0.2 nM [<sup>3</sup>H]NMS (a concentration that approximates the dissociation constant ( $K_d$ )) in the presence of various concentrations of C<sub>7</sub>/3-phth, gallamine, alcuronium or 4P-C<sub>7</sub>/3-phth at 22 °C.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 December 2012; accepted 15 August 2013.

Published online 13 October 2013.

- Conn, P. J., Jones, C. K. & Lindsley, C. W. Subtype selective allosteric modulators of muscarinic receptors for the treatment of CNS disorders. *Trends Pharmacol. Sci.* **30**, 148–155 (2009).
- Keov, P., Sexton, P. M. & Christopoulos, A. Allosteric modulation of G protein-coupled receptors: a pharmacological perspective. *Neuropharmacology* **60**, 24–35 (2011).
- Filmore, D. It's a GPCR world. *Modern Drug Discov.* **7**, 24–28 (2004).
- Jakubik, J. & El-Fakahany, E. E. Allosteric modulation of muscarinic acetylcholine receptors. *Pharmaceuticals* **3**, 2838–2860 (2010).
- Haga, K. *et al.* Structure of human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* **482**, 547–551 (2012).

6. Liu, W. *et al.* Structural basis for allosteric regulation of GPCRs by sodium ions. *Science* **6091**, 232–236 (2012).
7. Rasmussen, S. G. *et al.* Crystal structure of the  $\beta_2$  adrenergic receptor–Gs protein complex. *Nature* **477**, 549–555 (2011).
8. Lazareno, S. & Birdsall, N. J. Detection, quantitation, and verification of allosteric interactions of agents with labeled and unlabeled ligands at G protein-coupled receptors: interactions of strychnine and acetylcholine at muscarinic receptors. *Mol. Pharmacol.* **48**, 362–378 (1995).
9. Dror, R. O. *et al.* Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl Acad. Sci. USA* **108**, 13118–13123 (2011).
10. Shan, Y. *et al.* How does a drug molecule find its target binding site? *J. Am. Chem. Soc.* **133**, 9181–9183 (2011).
11. Buch, I., Giorgino, T. & De Fabritiis, G. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl Acad. Sci. USA* **108**, 10184–10189 (2011).
12. Prilla, S., Schrobang, J., Ellis, J., Hölte, H. D. & Mohr, K. Allosteric interactions with muscarinic acetylcholine receptors: complex role of the conserved tryptophan M<sub>2</sub><sup>422</sup>Trp in a cryptical cluster of amino acids for baseline affinity, subtype selectivity, and cooperativity. *Mol. Pharmacol.* **70**, 181–193 (2006).
13. Huang, X.-P., Prilla, S., Mohr, K. & Ellis, J. Critical amino acid residues of the common allosteric site on the M2 muscarinic acetylcholine receptor. *Mol. Pharmacol.* **68**, 769–778 (2005).
14. May, L. T. *et al.* Structure-function studies of allosteric agonism at M2 muscarinic acetylcholine receptors. *Mol. Pharmacol.* **72**, 463–476 (2007).
15. Trankle, C. *et al.* Interactions of orthosteric and allosteric ligands with [<sup>3</sup>H]dimethyl-W84 at the common allosteric site of muscarinic M2 receptors. *Mol. Pharmacol.* **64**, 180–190 (2003).
16. Ballesteros, J. & Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G-protein-coupled receptors. *Methods Neurosci.* **25**, 366–428 (1995).
17. Matsui, H., Lazareno, S. & Birdsall, N. J. Probing of the location of the allosteric site on m1 muscarinic receptors by site-directed mutagenesis. *Mol. Pharmacol.* **47**, 88–98 (1995).
18. Ma, L. *et al.* Selective activation of the M<sub>1</sub> muscarinic acetylcholine receptor achieved by allosteric potentiation. *Proc. Natl Acad. Sci. USA* **106**, 15950–15955 (2009).
19. Daiss, J. O. *et al.* N<sup>+</sup>/Si replacement as a tool for probing the pharmacophore of allosteric modulators of muscarinic M<sub>2</sub> receptors: synthesis, allosteric potency, and positive cooperativity of silicon-based W84 derivatives. *Organometallics* **21**, 803–811 (2002).
20. Choe, H. W. *et al.* Crystal structure of metarhodopsin II. *Nature* **471**, 651–655 (2011).
21. Bock, A. *et al.* The allosteric vestibule of a seven transmembrane helical receptor controls G-protein coupling. *Nat. Commun.* **3**, 1044 (2012).
22. Shoichet, B. & Kobilka, B. Structure-based drug screening for G-protein-coupled receptors. *Trends Pharmacol. Sci.* **33**, 268–272 (2012).
23. Totrov, M. & Abagyan, R. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr. Opin. Struct. Biol.* **18**, 178–184 (2008).
24. Avlani, V., May, L. T., Sexton, P. M. & Christopoulos, A. Application of a kinetic model to the apparently complex behavior of negative and positive allosteric modulators of muscarinic acetylcholine receptors. *J. Pharmacol. Exp. Ther.* **308**, 1062–1072 (2004).
25. Gao, Z.-G. *et al.* Identification of essential residues involved in the allosteric modulation of the human A<sub>3</sub> adenosine receptor. *Mol. Pharmacol.* **63**, 1021–1031 (2003).
26. Silvano, E. *et al.* The tetrahydroisoquinoline derivative SB269,652 is an allosteric antagonist at dopamine D3 and D2 receptors. *Mol. Pharmacol.* **78**, 925–934 (2010).
27. Lazareno, S., Popham, A. & Birdsall, N. J. Analogs of WIN 62,577 define a second allosteric site on muscarinic receptors. *Mol. Pharmacol.* **62**, 1492–1505 (2002).
28. Yanamala, N. & Klein-Seetharaman, J. Allosteric modulation of G protein coupled receptors by cytoplasmic, transmembrane, and extracellular ligands. *Pharmaceuticals* **3**, 3324–3342 (2010).
29. Shaw, D. E. *et al.* Millisecond-scale molecular dynamics simulation on Anton. In *Proceedings of the Conference on High Performance Computing, Networking, Storage, and Analysis* (ACM Press, 2009); available at <http://dl.acm.org/citation.cfm?id=1654099> (2009).
30. MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank T. Mildorf, A. Kruse and B. Kobilka for comments; J. Klepeis, B. Gregersen, J.-L. Li, K. Palmo, A. Donchev and particularly A. Taube for advice and support related to force fields and quantum mechanical calculations; Z. Fan for assistance with statistical analysis; A. Lerer and T. O'Donnell for assistance with simulation and analysis software; A. Philippsen for creating the video; A. Stewart for assistance with mutagenesis and cell line generation; K. Ban and J. Harjani for assistance with chemical synthesis; J. Swarbrick for recording and analysing the two-dimensional NMR data; B. Sleebs and S. Marcuccio for provision of synthetic reagents; J. Dang for advice on analytical chemistry; and M. Kirk and R. Kastleman for editorial assistance. Portions of this work were financed by Program Grant no. 519461 from the National Health and Medical Research Council (NHMRC) of Australia, with synthetic chemistry infrastructure support from the Australian Federal Education Investment Fund Super Science Initiative and Victoria's Science Agenda Investment Fund. A.C. and P.M.S. are Principal Research Fellows of the NHMRC; J.B.B. is a Senior Research Fellow of the NHMRC; J.R.L. is a Career Development Awardee of the NHMRC.

**Author Contributions** R.O.D. conceived this study and, with D.E.S., oversaw molecular dynamics simulations and analysis. R.O.D., H.F.G., D.W.B., J.R.V., A.C.P. and D.H.A. designed and analysed molecular dynamics simulations. H.F.G., J.R.V., A.C.P. and D.H.A. performed molecular dynamics simulations. R.O.D., H.F.G., D.W.B. and J.R.V. performed computational design of receptor mutants and of the modulator 4P-C<sub>7</sub>/3-phth. C.V. performed all biological assays and, with J.R.V. and A.C., analysed experimental data. M.C. and J.R.L. performed mutagenesis and generated the stable cell lines. J.B.B. and R.R. designed, planned and executed the synthesis of 4P-C<sub>7</sub>/3-phth, with active input from D.W.B. P.M.S. and A.C. supervised the cell-based biological studies. R.O.D., H.F.G., D.W.B., A.C. and D.E.S. wrote the manuscript. R.O.D., A.C. and D.E.S. supervised the overall research.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.O.D. (Ron.Dror@DEShawResearch.com), A.C. (Arthur.Christopoulos@Monash.edu) or D.E.S. (David.Shaw@DEShawResearch.com).

## METHODS

**System setup for molecular dynamics simulations.** Simulations of the human M2 muscarinic acetylcholine receptor (M2 receptor) were based on the M2–QNB complex crystal structure (Protein Data Bank accession number 3UON; ref. 5). Coordinates were prepared by removing the T4 lysozyme insert and capping residues 217 and 377 with *N*-methyl amide and acetyl groups, respectively (that is, intracellular loop 3 was absent from the simulations). Side chains for Lys 383<sup>6,31</sup> and Lys 384<sup>6,32</sup>, not resolved in the crystal structure, were modelled by hand. Hydrogen atoms were added using Maestro (Schrödinger), as previously described<sup>31</sup>. Histidine residues were simulated as the neutral *N*<sub>ε</sub> tautomer. Other titratable residues were simulated in their dominant protonation state at pH 7 except for Asp 69<sup>2,50</sup>, which was protonated; Asp<sup>2,50</sup> is protonated during the entire photocycle in rhodopsin<sup>32</sup>. We performed a control simulation of C<sub>7</sub>/3-phth binding to an unliganded M2 receptor in which Asp 69<sup>2,50</sup> was deprotonated (Supplementary Table 2, condition AO); a sodium ion bound to D69<sup>2,50</sup>, and afterwards, C<sub>7</sub>/3-phth bound in a pose that closely matched the pose seen in our other simulations, in which D69<sup>2,50</sup> was protonated. Ligand protonation states are indicated in Supplementary Fig. 1. In particular, strychnine was simulated in the protonated form (pK<sub>a</sub> 8.26; ref. 33).

The M2 receptor was inserted into an equilibrated palmitoylcholine bilayer as described previously<sup>34</sup>. Sodium and chloride ions were added to neutralize the system, and extra NaCl was added to reach a final concentration of approximately 150 mM. In each simulation, the simulation box initially measured 79 Å × 79 Å × 85 Å, and contained one M2 receptor, approximately 156 palmitoylcholine molecules, approximately 24 sodium ions, approximately 35 chloride ions and approximately 9,165 water molecules, for a total of approximately 53,000 atoms. A control simulation of C<sub>7</sub>/3-phth binding to an unliganded M2 receptor was designed to match radioligand binding experimental conditions closely; it contained approximately 10 mM MgCl<sub>2</sub> in addition to 100 mM NaCl (Supplementary Table 2, condition AP). This simulation (run at 22 °C rather than the usual 37 °C; see below) showed no noticeable differences in behaviour from our other simulations of C<sub>7</sub>/3-phth binding; in particular, C<sub>7</sub>/3-phth adopted the same bound pose.

Depending on the simulation, the orthosteric site was either unoccupied or occupied by QNB or NMS. For simulations with QNB in the orthosteric site, the coordinates from the M2–QNB complex were used<sup>5</sup>. For simulations with an unoccupied orthosteric site, the co-crystallized QNB was deleted from the protein. For simulations with NMS in the orthosteric site, the M3 muscarinic receptor–tiotropium complex crystal structure (Protein Data Bank accession number 4DAJ; ref. 35) was used to model the bound pose of NMS. The receptors were aligned using the C $\alpha$  atoms of residues within 5 Å of tiotropium, and tiotropium was morphed into NMS; the NMS phenyl group was positioned to overlap the more intracellular of the two tiotropium thienyl groups (Supplementary Fig. 23).

**Simulation protocol.** All simulations were performed using Anton, a special-purpose computer designed to accelerate standard molecular dynamics simulations by orders of magnitude<sup>29,36</sup>. Simulations were equilibrated in the NPT ensemble (37 °C (310 K), 1 bar (~1 atm)), with 5 kcal mol<sup>-1</sup> Å<sup>-2</sup> harmonic position restraints applied to all non-hydrogen protein and ligand atoms; restraints were tapered off linearly over 50 ns. Bond lengths to hydrogen atoms were constrained using M-SHAKE<sup>37</sup>, and an r-RESPA integrator<sup>38</sup> was used (2-fs time step; long-range electrostatics computed every 6 fs). Production simulations used the same integrator, pressure and temperature, and were initiated from the final snapshot of the corresponding equilibration simulation, with velocities sampled from the Boltzmann distribution at 310 K. Van der Waals and short-range electrostatic interactions were cut off at 13.5 Å, and long-range electrostatic interactions were computed using the *k*-space Gaussian split Ewald method<sup>39</sup> with a 32 × 32 × 32 grid,  $\sigma = 3.33$  Å and  $\sigma_s = 2.33$  Å.

**Spontaneous binding of allosteric modulators.** In simulations of known allosteric modulators (C<sub>7</sub>/3-phth [heptane-1,7-bis(dimethyl-3'-phthalimidopropylammonium)], dimethyl-W84 [hexane-1,6-bis(dimethyl-3'-(4-methyl-phthalimido)propyl ammonium)], gallamine, strychnine, alcuronium, or hexamethonium), the new designed modulator 4P-C<sub>7</sub>/3-phth [heptane-1,7-bis(dipropyl-3'-phthalimidopropyl ammonium)], or hypothetical modulators (4I-C<sub>7</sub>/3-phth [heptane-1,7-bis(diisopropyl-3'-phthalimidopropylammonium)] or 4C-C<sub>7</sub>/3-phth [heptane-1,7-bis(dicyclopentyl-3'-phthalimidopropylammonium)]), one modulator molecule was placed at an arbitrary location in the aqueous region, more than 20 Å from the centre of the extracellular vestibule, and allowed to diffuse freely until it associated spontaneously with the receptor, as we described previously for  $\beta$ -adrenergic receptors<sup>9</sup>. No artificial forces were applied to the ligand. Initial conformations of strychnine and alcuronium were based on crystal structures of strychnine and dimethyl caracurine-II (ref. 40), respectively.

**Simulation of tetraethylammonium.** To probe the relative importance of centre 1 and centre 2 in ligand binding, we used tetraethylammonium (TEA), a simple

molecule containing a single quaternary ammonium group. We performed a 5.0- $\mu$ s simulation of one TEA molecule binding spontaneously to an unliganded M2 receptor, following the protocol described above (Supplementary Table 2, condition AN). TEA bound and dissociated rapidly, allowing estimation of its relative affinity for the two centres (Supplementary Fig. 13).

**Design and simulation of mutated receptors.** Before performing the radioligand binding experiments described below (see section 'Pharmacological assays'), we performed simulations of C<sub>7</sub>/3-phth binding to several M2 receptor mutants we had designed—on the basis of wild-type receptor simulations—to increase or decrease allosteric modulator binding affinity. These simulations (conditions N–AB; see Supplementary Table 2), were performed as described above. The mutated residues were modelled using the PyMOL mutagenesis wizard<sup>41</sup>. In simulations containing N419W, N410Y, T84D and/or T423E mutations (conditions N–P and Z)—mutations designed to increase modulator affinity—C<sub>7</sub>/3-phth adopted the same pose as it did when binding to the wild-type receptor, and made the predicted favourable interactions (cation– $\pi$  or cation–anion) with the mutated residues. In the simulation of the N410K + T423K mutant (condition Q)—designed to decrease modulator affinity—C<sub>7</sub>/3-phth never associated with the receptor. In simulations of the Y80K and Y83K mutants (conditions AA and AB), C<sub>7</sub>/3-phth associated with the receptor only transiently, never reaching a stable bound conformation. Simulations of several other mutants designed to increase modulator affinity yielded inconclusive results, typically because C<sub>7</sub>/3-phth adopted a different binding pose than observed in wild-type simulations (conditions R–V). We thus chose to express and perform radioligand binding experiments on those mutants shown in Fig. 2.

To probe ligand–receptor interactions at centre 2 further, we also performed simulations of C<sub>7</sub>/3-phth binding to the Y80A, L82A and Y83A mutants (conditions W–Y, respectively; cf. Supplementary Fig. 14).

**Design and simulation of C<sub>7</sub>/3-phth analogues.** We sought to design a C<sub>7</sub>/3-phth analogue that would maintain the same binding pose but require a wider conformation of the extracellular vestibule. For each of several candidate analogues, we performed *de novo* binding simulations to an NMS-bound M2 receptor (Supplementary Table 2, conditions AH, AK, AL), to test whether the candidate would bind in a pose analogous to that of C<sub>7</sub>/3-phth. (These were sometimes supplemented by simulations of the NMS-occupied receptor with the candidate initially modelled into the extracellular vestibule in the C<sub>7</sub>/3-phth pose (Supplementary Table 2, conditions AF, AI).) We also performed simulations of the NMS-free receptor with each candidate initially modelled into the extracellular vestibule in the C<sub>7</sub>/3-phth pose (Supplementary Table 2, conditions AG, AJ, AM), to test whether the candidate would maintain a wide vestibule conformation. 4P-C<sub>7</sub>/3-phth satisfied all criteria (Fig. 3d and Supplementary Fig. 19) and was thus selected for synthesis and experimental evaluation. Other candidates failed our computational tests, in that they either bound to the NMS-free receptor without widening the vestibule or failed to bind in a pose similar to that of C<sub>7</sub>/3-phth.

**Force field parameters.** We used the CHARMM27 parameter set for protein molecules and salt ions and the CHARMM TIP3P model for water<sup>30</sup>; protein parameters incorporated CMAP terms<sup>42</sup> and modified Asp, Glu and Arg side-chain partial charges<sup>43</sup>. We used a modified CHARMM lipid force field<sup>44</sup>. Ligand parameters were based on results from the CHARMM ParamChem web server, version 0.9.6  $\beta$  (ref. 45), and incorporated specific modifications to partial charges, dihedral terms and angle terms, as shown in Supplementary Fig. 24. When ParamChem reported large penalties (errors) for its estimated partial charges, better estimates were determined by submitting appropriately chosen fragments or analogues to ParamChem (Supplementary Fig. 24). When large penalties were reported for a dihedral term, a new term was created by fitting to high-level DF-LMP2/aug-cc-pVTZ *ab initio* quantum mechanical calculations from a locally tuned version of MOLPRO version 2010.1 (H.-J. Werner *et al.*, Cardiff University, 2010). Force field parameters for all ligands are available upon request.

Because commonly used molecular dynamics force fields are known to model cation– $\pi$  interactions poorly<sup>46</sup>, we modified certain force field terms to better represent cation– $\pi$  interactions involving simulated ligands. We began by characterizing the error that results from using standard force field parameters to determine the aqueous benzene–tetramethylammonium (TMA) association constant. We simulated one benzene molecule and two TMA molecules in 150 mM NaCl for 300 ns. Benzene was parameterized using the CHARMM phenylalanine parameters; TMA was parameterized using the ParamChem server. The benzene–TMA association constant (*K*<sub>A</sub>) was determined on the basis of the fraction of time in which the benzene formed a complex with either TMA. Although a simple model based on experimental data predicts a *K*<sub>A</sub> of about 1.8 M<sup>-1</sup> (ref. 47), our simulation yielded a value of only 0.67 M<sup>-1</sup>, corresponding to a binding energy discrepancy of approximately 0.6 kcal mol<sup>-1</sup> (standard force field parameters led to weaker association than is observed experimentally). We strengthened benzene–TMA association by increasing the  $\epsilon$  parameter for the van der Waals (Lennard-Jones)



interactions between carbon atoms of benzene and carbon atoms of TMA from 0.0734 kcal mol<sup>-1</sup> to 0.175 kcal mol<sup>-1</sup> (the  $\sigma$  parameter, 3.8 Å, was not changed). Simulations with these modified parameters yielded a more realistic  $K_A$  of 1.86 M<sup>-1</sup>. We used this modified C–C  $\epsilon$  parameter for van der Waals interactions between those ligand carbon atoms directly bonded to quaternary ammonium nitrogen atoms and the carbon atoms of all aromatic rings in the M2 receptor.

Because strychnine has a tertiary, not quaternary, ammonium group, we also developed an alternative correction involving modification of the van der Waals interaction between the ligand ammonium nitrogen atom and carbon atoms of aromatic residues. In this case, we modified the  $\epsilon$  parameter from 0.118 kcal mol<sup>-1</sup> to 0.385 kcal mol<sup>-1</sup> (the  $\sigma$  parameter, 3.4 Å, was not changed). Simulations with these modified parameters yielded a benzene–TMA  $K_A$  of 1.83 M<sup>-1</sup>. We used this C–N  $\epsilon$  parameter modification in simulations of strychnine. As a control, we also simulated *N*-methylstrychnine (which has a quaternary ammonium group but has properties similar to those of strychnine) with the C–C  $\epsilon$  parameter modification and verified that its behaviour in simulation was similar to that of the strychnine simulations. Likewise, we confirmed that C<sub>7</sub>/3-phth and dimethyl-W84 simulations using the C–C  $\epsilon$  parameter modification (condition A, simulation 5; condition B, simulation 5; condition J, simulation 3; condition K, simulation 3; see Supplementary Table 2) yielded behaviour matching that of simulations using the C–N  $\epsilon$  parameter modification (all other simulations under conditions A, B, J and K). We used the C–N  $\epsilon$  parameter modification in conditions C, D, L–V and AE (see Supplementary Table 2), and the C–C  $\epsilon$  parameter modification in all other simulations.

**Analysis protocols.** Trajectory frames (atomic coordinates) were saved every 180 ps during production simulations. VMD was used to visualize trajectories<sup>48</sup>, and molecular images were rendered using PyMOL<sup>41</sup> and POV-Ray (Persistence of Vision; <http://www.povray.org>). The ammonium group locations shown in Fig. 1e were sampled at 1-ns intervals from 1.0- $\mu$ s simulation windows. Cross sections in Fig. 3c, d are on a plane through Y177<sup>ECL2</sup> and N419<sup>7.32</sup> perpendicular to the lipid bilayer.

The root mean squared deviation (r.m.s.d.) of ligand atomic positions was calculated by first aligning the receptor transmembrane helices (C $\alpha$  atoms of residues 20–49, 57–86, 93–126, 138–166, 184–213, 383–412 and 419–439), and then computing the r.m.s.d. of non-hydrogen ligand atoms. For each ligand, the binding pose adopted in most simulations was declared the consensus pose. A single representative of this consensus pose was used to compute r.m.s.d.; this representative pose is shown in all figures, unless otherwise indicated. All r.m.s.d. traces were smoothed using a 10-ns running average.

An estimate of the energetic contribution of electrostatic repulsion to cooperativity—the difference between the allosteric modulator's electrostatic binding energy in the presence and absence of NMS—was determined as the sum, over all modulator atoms, of the atomic partial charge multiplied by the receptor electrostatic difference potential (NMS-bound minus unliganded) evaluated at the position of each modulator atom. Electrostatic potentials were determined by solving the linearized Poisson–Boltzmann equation at an ionic strength of 150 mM using APBS<sup>49</sup>; only protein and (when present) NMS atoms were included in this calculation. For each allosteric modulator, this calculation was performed for 10 frames evenly spaced over that portion of a trajectory during which the modulator was bound. The average of these sums is the electrostatic energetic contribution for that modulator in that particular trajectory, and each of the values in Supplementary Table 4 represents a further average over several trajectories. For the PAMs (alcuronium and strychnine), we used trajectories with NMS bound in the orthosteric site; NMS was removed before ABPS calculation of the 'unliganded' electrostatic potentials. For the NAMs (C<sub>7</sub>/3-phth, dimethyl-W84 and gallamine), we used trajectories in which the orthosteric site was unoccupied; before ABPS calculation of the 'NMS-bound' electrostatic potentials, NMS was docked manually into the orthosteric site, and receptor atomic positions were minimized to remove steric clashes.

**Docking protocol.** For purposes of comparison, we performed flexible ligand docking using Glide (Schrödinger), using the same human M2 receptor structure as the modulator binding molecular dynamics simulations described above. Modulators were allowed to dock in a 25-Å box centred on the centroid of residues E172<sup>ECL2</sup>–E175<sup>ECL2</sup>, Y177<sup>ECL2</sup>, N419<sup>7.32</sup>, W422<sup>7.35</sup> and T423<sup>7.36</sup>, which were selected on the basis of mutagenesis data suggesting the general location of the allosteric binding site<sup>12–14,50,51</sup>. This docking box contained, and indeed was roughly centred on, the binding poses observed in molecular dynamics simulations. The results achieved with a 25-Å box did not differ qualitatively from the results of control docking experiments using a 15-Å box. Each ligand was docked in Glide, using standard-precision docking with the default settings. Ten docked poses for each ligand were obtained, and these docking predictions were compared with the pose obtained by molecular dynamics (Supplementary Fig. 21). Control docking experiments with extra-precision (XP) settings did not produce results more similar to those obtained by molecular dynamics simulation. This same procedure was also used to dock modulators to receptor coordinates taken from

a molecular dynamics simulation of the unliganded M2 receptor (condition AD; see Supplementary Table 2).

**Pharmacological assays.** Reagents were obtained from the following vendors: Dulbecco's modified Eagle's medium (DMEM), hygromycin-B and CHO-FlpIn cells (Invitrogen); fetal bovine serum (FBS; Thermo Electron); [<sup>3</sup>H]NMS (>82.0 Ci mmol<sup>-1</sup>; PerkinElmer Life Sciences); oligonucleotides (Geneworks). Mycoplasma testing was performed regularly on cell lines using the MycoAlert<sup>TM</sup> kit (Lonza); cell lines were mycoplasma-free before and after experiments were conducted. All other reagents, purchased from Sigma, were analytical grade.

**Mutagenesis and stable cell line generation.** Site-directed mutagenesis of the human M2 receptor was performed in the pENTR/D-TOPO vector before transfer to the pEFS/FRT/V5-DEST vector, as previously described<sup>14,52</sup>, to produce a receptor with a V5 epitope tag at the carboxy (C) terminus. The V5-tagged wild-type receptor has been shown to possess pharmacological properties equivalent to those of the untagged receptor<sup>24</sup>. Receptor constructs were integrated isogenically into CHO-FlpIn cells, as previously described<sup>14</sup>. Stably transfected cell lines were maintained in DMEM supplemented with 10% FBS and 600  $\mu$ g ml<sup>-1</sup> hygromycin-B ('complete DMEM') at 37 °C in a humidified incubator containing 5% CO<sub>2</sub>, 95% O<sub>2</sub>. A list of mutagenesis primers used can be found in Supplementary Table 8.

**Radioligand binding assays.** Radioligand binding experiments were performed on CHO-FlpIn whole cells stably expressing the human M2 receptor construct of choice. After plating 20,000 cells in complete DMEM into 96-well ISOPATE TC plates (all amounts are per well), cells were allowed to grow overnight at 37 °C. The next day, cells were washed with phosphate-buffered saline (100  $\mu$ l) and resuspended in binding buffer (10 mM HEPES, 100 mM NaCl, 10 mM MgCl<sub>2</sub>, pH 7.4). Assay mixtures, in a total volume of 200  $\mu$ l with a 1/10 dilution of drug, were incubated at room temperature (22 °C) for 4 h, except for the comparison between C<sub>7</sub>/3-phth, 4P-C<sub>7</sub>/3-phth and alcuronium at the wild-type receptor, which was performed over 24 h. Assays were terminated by buffer removal followed by rapid washing, twice, with ice-cold 0.9% NaCl (100  $\mu$ l). Plates were allowed to dry inverted for 30 min; OptiPhase Supermix scintillation cocktail (100  $\mu$ l) was added, plates were sealed (TopSeal<sup>TM</sup>) and radioactivity was measured in a MicroBeta<sup>2</sup> LumijET microplate counter. Saturation binding experiments were performed in the absence or presence of atropine (100  $\mu$ M) with 0.003–3 nM [<sup>3</sup>H]NMS. Inhibition binding experiments were performed with 0.2 nM [<sup>3</sup>H]NMS (the approximate  $K_d$ ) in the presence of various concentrations of C<sub>7</sub>/3-phth, gallamine, 4P-C<sub>7</sub>/3-phth or alcuronium. For [<sup>3</sup>H]NMS dissociation kinetics experiments, cells were first incubated with the radioligand at room temperature (22 °C) for 4 h before initiating radioligand dissociation, using atropine (100  $\mu$ M) either in absence or presence of a high concentration (100  $\mu$ M) of a modulator (either C<sub>7</sub>/3-phth or gallamine).

The number of replicates in each group was chosen according to standard practices in the field. All experiments used  $n = 3$ , except for certain measurements of [<sup>3</sup>H]NMS dissociation rate and allosteric modulator affinity and cooperativity, which used  $n = 4$  or  $n = 5$  (see Supplementary Fig. 11 and Supplementary Tables 6 and 7).

As a control, we repeated several experiments in a buffer that did not contain magnesium ions. In particular, we removed MgCl<sub>2</sub> from the buffer described above and performed inhibition binding experiments for C<sub>7</sub>/3-phth at the wild-type, N410Y, T84D, N410Y + T423E and N410K receptors. All C<sub>7</sub>/3-phth affinity values increased by a factor of 2–4 in the absence of magnesium, but each mutation still led to approximately the same change in C<sub>7</sub>/3-phth affinity (data not shown).

**Data analysis.** Nonlinear regression was performed using Prism 5.04 (GraphPad Software). Total and non-specific [<sup>3</sup>H]NMS binding data were globally fitted to a one-site saturation binding model to derive estimates of the radioligand dissociation constant ( $K_d$ ) and the maximal density of binding sites ( $B_{max}$ ) for the wild-type receptor and each mutant (Supplementary Fig. 10). Dissociation kinetics data for all mutant receptors were fitted to a monoexponential function to derive observed dissociation rate constants ( $k_{off}$ ; Supplementary Fig. 11). With two exceptions (noted below), all interaction binding data were fitted to a kinetic allosteric ternary complex model (kinetic ATCM) to accommodate any potential non-equilibrium binding under our experimental conditions. For this latter analysis, the equations are as derived previously by Lazareno and Birdsall<sup>8</sup>, with the only modification being that the affinity constants in the original equations were re-cast as equilibrium dissociation constants<sup>24</sup>:

$$B_t = B_{AB} [1 - e^{-k_{on,obs}t}] \quad (1)$$

where

$$k_{on,obs} = k_{off,obs} (1 + [A]/K_{App}) \quad (2)$$

$$K_{App} = K_d \left( \frac{1 + [B]/K_i}{1 + \alpha[B]/K_i} \right) \quad (3)$$

$$k_{\text{off,obs}} = \frac{k_{\text{off}} + [B]k_{\text{off,B}}/(K_i/\alpha)}{1 + [B]/(K_i/\alpha)} \quad (4)$$

and

$$B_{\text{AB}} = \frac{[A]/K_{\text{App}}}{1 + [A]/K_{\text{App}}} \quad (5)$$

In these equations,  $[A]$  denotes the concentration and  $K_d$  the equilibrium dissociation constant of the orthosteric radioligand;  $[B]$  denotes the concentration and  $K_i$  the equilibrium dissociation constant for the allosteric modulator;  $\alpha$  is the cooperativity factor for allosteric interaction between radioligand and modulator;  $K_{\text{App}}$  denotes the apparent equilibrium dissociation constant of the orthosteric radioligand in the presence of modulator;  $B_t$  denotes the fractional binding of the radioligand at time  $t$ ;  $B_{\text{AB}}$  denotes the fractional binding of the radioligand in the presence of modulator at equilibrium;  $k_{\text{on,obs}}$  and  $k_{\text{off,obs}}$  denote the apparent association and dissociation rate constants for the radioligand in the presence of a modulator, respectively;  $k_{\text{off}}$  denotes the radioligand dissociation rate constant when the receptor is not occupied by modulator; and  $k_{\text{off,B}}$  denotes the radioligand dissociation rate constant for the modulator-occupied receptor; this latter value was not significantly different from zero, and was constrained as such. This analysis assumes that the binding kinetics of each modulator are rapid relative to the radioligand and that the modulator rapidly achieves equilibrium with the allosteric site, as is generally found for prototypical modulators of the muscarinic acetylcholine receptors<sup>8</sup>.

For the Y80K and N410K + T423K mutants, the affinity of the modulators was too weak to allow for a reliable fit of the full kinetic ATCM (the binding curves could not be fully defined). Thus, to obtain an estimate of the affinity of each modulator under these conditions, we fitted the model to each data set with the cooperativity parameter  $\alpha$  restrained to each of 11 values ranging from 1 to 0 in 0.1-unit increments. All fixed values of  $\alpha$  yielded similar affinity estimates: estimated  $pK_i$  ranged from 4.64 to 5.06 for C<sub>7</sub>/3-phth at Y80K, from 3.12 to 3.52 for gallamine at Y80K, from 3.88 to 4.21 for C<sub>7</sub>/3-phth at N410K + T423K and from 3.32 to 3.64 for gallamine at N410K + T423K. Thus, although we cannot estimate  $\alpha$  with confidence for these mutants, we can derive reasonably accurate estimates of affinity. In each case, the value of  $\alpha$  yielding the minimum mean squared error for each curve fit (0.5 and 0.7 for C<sub>7</sub>/3-phth at Y80K and N410K + T423K, respectively; 0.4 and 0.3 for gallamine at Y80K and N410K + T423K, respectively) was used to determine the estimate of modulator affinity shown in Fig. 2d. The error bars for Y80K and N410K + T423K represent a sum of the inherent error of the  $pK_i$  estimate with this fixed  $\alpha$  value and the change in  $pK_i$  estimates that resulted from setting  $\alpha$  to the most extreme values.

To determine statistical significance of changes in allosteric modulator affinity ( $pK_i$ ) values for mutant receptors, we used a standard one-way analysis of variance (ANOVA) test with an unprotected Fisher's least significant difference post-test; this test was one-tailed in cases where we predicted an affinity change in a specific direction (Fig. 2d) and two-tailed otherwise (Supplementary Fig. 14). We used the same one-tailed test to determine statistical significance of changes in cooperativity ( $\alpha$ ) for different modulators at the wild-type M<sub>2</sub> receptor (Fig. 3d). To determine the significance of changes in the affinity ( $pK_i$ ) of those different modulators to the wild-type M<sub>2</sub> receptor and of changes in the  $\alpha$ ,  $B_{\text{max}}$  and [<sup>3</sup>H]NMS  $pK_d$  values for mutant receptors (cases where we had no previous prediction of a change or of directionality of change), we used ANOVA with a two-tailed Dunnett's post-test (Supplementary Figs 10, 12 and 20). The statistical tests used make standard assumptions of data normality and equality of variance. The data did not show evidence of significant deviation from a normal distribution, and F-tests did not reject the hypothesis of equality of variance for all groups.

**Synthesis of the allosteric modulator 4P-C<sub>7</sub>/3-phth.** The allosteric modulator 4P-C<sub>7</sub>/3-phth was prepared in three steps, as follows. (1) Pimelic acid was activated with oxalyl chloride; the resulting acid chloride was then coupled with *N,N*-dipropylamine to afford *N,N,N',N'*-tetrapropyl heptanediamide. (2) The diamide was reduced with borane/THF or LiAlH<sub>4</sub> to afford *N,N,N',N'*-tetrapropyl heptane-1,7-diamine. (3) The diamine was quaternized with 2-(3-bromopropyl)-isoindoline-1,3-dione (ethylene carbonate solvent, sealed tube, 90 °C, 60 h, under N<sub>2</sub>) to give,

after HPLC purification, the target compound '4P-C<sub>7</sub>/3-phth', *N,N,N',N'*-bis-(3-(1,3-dioxisoindolin-2-yl)propyl)-*N,N,N',N'*-tetrapropyl heptane-1,7-diaminium bis-trifluoroacetate tetrahydrate. Additional detail on compound synthesis and characterization is provided in the Supplementary Information ('Materials and methods for organic synthesis' and Supplementary Fig. 25).

31. Dror, R. O. *et al.* Identification of two distinct inactive conformations of the  $\beta_2$ -adrenergic receptor reconciles structural and biochemical observations. *Proc. Natl Acad. Sci. USA* **106**, 4689–4694 (2009).
32. Fahmy, K. *et al.* Protonation states of membrane-embedded carboxylic acid groups in rhodopsin and metarhodopsin II: a Fourier-transform infrared spectroscopy study of site-directed mutants. *Proc. Natl Acad. Sci. USA* **90**, 10206–10210 (1993).
33. Everett, A. J., Openshaw, H. T. & Smith, G. F. The constitution of aspidospermine. Part III. Reactivity at the nitrogen atoms, and biogenetic considerations. *J. Chem. Soc.* 1120–1123 (1957).
34. Rosenbaum, D. M. *et al.* Structure and function of an irreversible agonist- $\beta_2$  adrenoceptor complex. *Nature* **469**, 236–240 (2011).
35. Kruse, A. *et al.* Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature* **482**, 552–556 (2012).
36. Shaw, D. E. *et al.* Atomic-level characterization of the structural dynamics of proteins. *Science* **330**, 341–346 (2010).
37. Kräutler, V., van Gunsteren, W. F. & Hünenberger, P. H. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J. Comput. Chem.* **22**, 501–508 (2001).
38. Tuckerman, M., Berne, B. J. & Martyna, G. J. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* **97**, 1990–2001 (1992).
39. Shan, Y., Klepeis, J. L., Eastwood, M. P., Dror, R. O. & Shaw, D. E. Gaussian split Ewald: a fast Ewald mesh method for molecular simulation. *J. Chem. Phys.* **122**, 54101 (2005).
40. Bourne, P. E., Ginell, S. L., Low, B. W. & Lessinger, L. Structure of a potent neuromuscular blocking agent: caracurine-II dimethochloride octahydrate, [C<sub>40</sub>H<sub>44</sub>N<sub>4</sub>O<sub>2</sub>]<sup>2+</sup>·2Cl<sup>−</sup>·8H<sub>2</sub>O. *J. Cryst. Spectroscop. Res.* **15**, 453–471 (1985).
41. DeLano, W. L. The PyMOL Molecular Graphics System v. 1.5.0.3-01 (Schrödinger, LLC, New York, New York, 2012).
42. Mackerell, A. D. Jr, Feig, M. & Brooks, C. L. III. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25**, 1400–1415 (2004).
43. Piana, S., Lindorff-Larsen, K. & Shaw, D. E. How robust are protein folding simulations with respect to force field parameterization? *Biophys. J.* **100**, L47–L49 (2011).
44. Klauda, J. B. *et al.* Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* **114**, 7830–7843 (2010).
45. Vanommeslaeghe, K. *et al.* CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–690 (2010).
46. Caldwell, J. & Kollman, P. Cation- $\pi$  interactions: nonadditive effects are critical in their accurate representation. *J. Am. Chem. Soc.* **117**, 4177–4178 (1995).
47. Schneider, H. *et al.* Host-guest supramolecular chemistry. 34. The incremental approach to noncovalent interactions: Coulomb and van der Waals effects in organic ion pairs. *J. Am. Chem. Soc.* **20**, 7698–7703 (1991).
48. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
49. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).
50. Gnagay, A. L., Seidenberg, M. & Ellis, J. Site-directed mutagenesis reveals two epitopes involved in the subtype selectivity of the allosteric interactions of gallamine at muscarinic acetylcholine receptors. *Mol. Pharmacol.* **56**, 1245–1253 (1999).
51. Voigtländer, U. *et al.* Allosteric site on muscarinic acetylcholine receptors: identification of two amino acids in the muscarinic M<sub>2</sub> receptor that account entirely for the M<sub>2</sub>/M<sub>5</sub> subtype selectivities of some structurally diverse allosteric ligands in *N*-methylscopolamine-occupied receptors. *Mol. Pharmacol.* **64**, 21–31 (2003).
52. Avlani, V. A. *et al.* Critical role for the second extracellular loop in the binding of both orthosteric and allosteric G protein-coupled receptor ligands. *J. Biol. Chem.* **282**, 25677–25686 (2007).

## CORRIGENDUM

doi:10.1038/nature12795

### Corrigendum: Stereoconversion of tertiary alcohols to tertiary-alkyl isonitriles and amines

Sergey V. Pronin, Christopher A. Reiher & Ryan A. Shenvi

*Nature* **501**, 195–199 (2013); doi:10.1038/nature12472

In Fig. 3 of this Letter and on pages SI-24 to SI-28 and SI-84 to SI-93 of the Supplementary Information, the Me and n-Pr labels should be reversed in compounds 17 to 21. Corresponding errors also affect the precursor alcohols 42, 44, 46, 47 and 48 on pages SI-24 to SI-28. The conclusions of the paper are not affected by these errors, which have been corrected in the online HTML and PDF versions of this Letter.



# CAREERS

**COLUMN** Expediting visa approval helps to lure research talent **p.303**

**@NATUREJOBS** Follow us on Twitter for the latest news and features [go.nature.com/e492gf](http://go.nature.com/e492gf)

**NATUREJOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)

D. INOUE



The Rocky Mountain Biological Laboratory field station outside Crested Butte, Colorado, offers high-tech facilities for fieldwork.

## FIELDWORK

# The great outdoors

*Field stations offer sophisticated facilities and opportunities for large-scale research.*

BY ROBERTA KWOK

Noah Whiteman's 2011 field season was tough. He and his team spent two summer months in the Rocky Mountains studying whether bacterial infections made plants more vulnerable to herbivores. They wanted to isolate bacteria from collected leaves to infect plants in the field, but the station at which they were working, the non-profit Rocky Mountain Biological Laboratory (RMBL) near Crested Butte, Colorado, did not have the equipment to support sterile laboratory work.

So two to three times every week, the team drove an hour each way to Western State Colorado University in Gunnison to autoclave nutrient media and pour it into Petri dishes

ready for growing bacteria. They were grateful for the facilities, but the process "was really cumbersome", says Whiteman, an ecological geneticist at the University of Arizona in Tucson. "We were exhausted."

But by Whiteman's 2012 field season, the RMBL had built a new research centre — with Bunsen burners, microfiltered water and fume cupboards for chemical work — mainly with funding from the 2009 stimulus package from the US federal government. The facilities got even better in 2013, when funding from the US National Science Foundation (NSF) and private donors allowed the station to add an autoclave, a shaking incubator, a polymerase-chain-reaction machine and a  $-80^{\circ}\text{C}$  freezer. The team could now store more plant and bacteria samples and process them much more quickly.

Many field stations used to offer biologists little more than access to the land, basic equipment such as microscopes and a place to sleep. But over the past decade or so, stations around the world have begun adding more sophisticated features: molecular-biology equipment, Wi-Fi, Global Positioning System (GPS) devices and features ranging from towers that allow researchers to monitor the forest canopy to facilities for conducting large-scale lake experiments. The upgrades, often funded by government grants, are driven partly by the falling cost of technology. Meanwhile, there is growing scientific interest in complex, large-scale research questions — including projects on the effects of climate change, invasive species and pathogens across entire regions. To support this work, programmes such as the ►

► NSF-funded US National Ecological Observatory Network (NEON), based in Boulder, Colorado, are collecting standardized ecological and atmospheric data across whole countries — a far cry from the simple collection of flora and fauna that once characterized field-station research. “Back in the day, you were just grabbing creatures,” says Sarah Oktay, director of the Nantucket Field Station operated by the University of Massachusetts Boston. Now, she says, scientists are more interested in big questions relevant to entire regions.

### HIGH-TECH FACILITIES

Upgraded Internet access at some stations allows researchers to upload data automatically from field sensors to university servers or data repositories. In 2006, with funding from the NSF, the Nantucket Field Station installed Wi-Fi over about 80% of its 43-hectare site, giving scientists online access to sensors to track species of interest (one team installed cameras to record videos of fiddler crabs) or monitor temperature, light or soil moisture. That speeds up access to data and makes it easier to find out whether a sensor has broken down. Otherwise Oktay might “have to put on a pair of waders and drive about 20 miles” to discover a broken sensor, she says.

Improved mapping technology allows scientists to collect precise location data. In 2004, the RMBL began offering handheld GPS units to visiting researchers. At the time, the units were able to pinpoint locations to within 100 centimetres; 2 years ago, with access to more satellite data, they could reach 20–40 centimetres. The station maintains a database of spatial information such as elevation data and locations of roads and species, so that researchers can look for correlations between their observations and other factors, for example between animals’ adrenaline levels and the creatures’ distance from recreational trails. And scientists conducting long-term studies can be confident that they are returning

to the same spots year after year, instead of relying on physical markers that might be moved or buried by wildlife.

Some field stations have installed especially sophisticated experimental facilities. In 2008, with NSF funding, the La Selva Biological Station in northern Costa Rica, managed by the non-profit Organization for Tropical Studies in Durham, North Carolina, completed 3 towers between 33 and 41 metres high that were equipped with electricity and Wi-Fi, allowing easy access to the forest canopy. The towers were constructed to make the station more attractive to researchers and educators, says station director Carlos de la Rosa. A robot shuttles between towers on cables and collects data on surface reflection, solar radiation and sound; weather stations gather meteorological data at various heights. Researchers can study differences between ecosystems on the ground and those high above the forest floor, such as variations in the types of beetle present. One team, says de la Rosa, is considering recording butterflies with cameras at different levels, then streaming the footage to the Internet. The researchers could then recruit members of the public to watch the videos and help to identify species, saving researchers many hours of viewing time.

### WATER WORK

Studies of aquatic ecosystems are also benefiting from better equipment. Last year, the Leibniz Institute of Freshwater Ecology and Inland Fisheries in Neuglobsow, Germany, set up its LakeLab facility to improve studies on the effects of climate change. Funded mainly by a grant from the Federal Ministry of Education and Research, the system consists of 24 cylindrical enclosures in Lake Stechlin, inside which scientists can simulate changes in the depth of the upper level of warm water, and measure the effects on flora and fauna. Researchers from Germany and Hungary are studying how these changes affect ciliates and algae, for example.

## INTO THE WILD

### *Planning is the key to success*

The basics of planning a successful visit to a field station have not changed. Be sure to book well in advance: some stations can fill up a year ahead. Consider whether the weather will be conducive to the work. “I’ve had people come and look for clams, and I’m like, well, they’re going to be under a foot of ice,” says Sarah Oktay, director of the Nantucket Field Station operated by the University of Massachusetts Boston.

Let the director know what equipment is needed and bring back-up electronics. If the station is not near civilization, don’t count on being able to buy extra batteries. And avoid

planning a packed schedule; researchers often underestimate how long fieldwork will take, and have to leave without enough data.

Once at the site, be open to opportunities and conversation, says Ian Billick, director of the Rocky Mountain Biological Laboratory near Crested Butte, Colorado. Visiting another scientist’s field site can foster rich discussions about the complexities of the surrounding environment and lead to new research directions — an experience sometimes lacking in the laboratory, where the focus tends to be limited to a few processes of interest. **R.K.**



Jörg Sareyka (left) and Mark Gessner collect zooplankton samples at LakeLab, Germany.

The enclosures are large — about 20 metres deep and 9 metres wide — reducing the influence of the walls and allowing experiments to better mimic natural conditions, says Mark Gessner, director of LakeLab.

The advantages offered by modern field stations go beyond technology: researchers stand to gain from many years of collective wisdom. Ecological studies that began in the 1960s and 1970s have laid the groundwork for a new generation of researchers, says Ian Billick, director of the RMBL. Today, field stations provide accumulated intellectual capital: decades of original data sets, archived research plans, specimen collections and oral history that can prove invaluable to young investigators. A researcher who wants to know the best place to find a particular flower species can easily get tips from a field-station staff member or a colleague who has worked at the site, instead of sifting through published papers. “That background information is critical for allowing new scientists to get projects up and running quickly,” says Billick (see ‘Planning is the key to success’).

### GOING BIG

Field-station directors often coordinate with each other and share data, allowing scientists to conduct large-scale studies. Over the past 10–20 years, the non-profit Organization of Biological Field Stations (OBFS) in Woodside, California, has worked for increased communication, says Oktay, who is secretary of the organization. Directors are also keen to support individual researchers. For example, if a scientist approaches the OBFS with an interest in studying a moth parasite in New England, the organization will e-mail all field-station



directors in the region to find out who has collected that parasite or moth, helping the researcher to decide where to visit.

Many stations are putting their data online, making it easier for scientists to spot trends. Black Rock Forest, a field station in Cornwall, New York, already has weather and other environmental sensor data posted on a project website and plans to add hydrology and tree-growth data, as well as all of its research publications, by next year, says Bill Schuster, the station's executive director.

Federally funded programmes are gathering data on larger scales, and making them accessible to all researchers. NEON aims to collect ecological observations at 106 sites across the United States for 30 years. It has installed basic infrastructure such as sensor towers and instrument huts at 24 sites, and hopes to complete all sites by 2017, at a cost of \$434 million. Sensors will monitor features such as atmospheric conditions and soil properties at terrestrial sites, and pH, oxygen and nutrient levels in streams; field crews will study organisms and fly instrumented planes to take high-resolution aerial images and survey vegetation. In Australia, the Terrestrial Ecosystem Research Network (TERN) based in St Lucia performs a similar function, gathering data on greenhouse-gas exchange, soil characteristics, bioacoustics and weather at ten sites across the country.

NEON and TERN each collect their data in a standardized way so that researchers can make comparisons on regional or continental scales. Without that consistency, "you literally cannot say what is happening across the country", says Suzanne Long, the executive knowledge broker at TERN, who works to increase the network's impact on environmental policy and practice in Australia.

The data could complement work at specific field sites, suggests Dave Tazik, director of biology at NEON; the observatory's sites could serve as control plots while researchers experimentally manipulate other plots. "We think of it as providing a backbone of data," he says. Scientists can also apply to add their own sensors to NEON towers. Researchers could develop a hypothesis on the basis of their field-station work — perhaps investigating how plants respond to climate change — and then analyse the freely available NEON data to explore whether trends hold true on a broad scale.

Now more than ever, field stations are helping scientists to tackle big questions about pressing environmental issues, such as the effect of climate change on ecosystems. "As these systems fall apart, it's actually a very good time to study them," says Billick. ■

**Roberta Kwok** is a freelance science writer in Seattle, Washington.

## COLUMN

# Fast-tracked talent

Expediting visa approval helps countries to attract the best researchers, says **Conor O'Carroll**.

Encouraging and facilitating mobility among scientists helps both the researcher and the country thirsty for talent. Indeed, the success of the €70.2-billion (US\$95-billion) Horizon 2020 European Union (EU) research programme, set to be launched on 11 December, hinges on bringing researchers to Europe from around the globe. And yet obtaining a visa, the key to being mobile, can be a major challenge. Ireland's experience with smoothing the visa process shows one way to make the system more efficient while maintaining border security.

The EU has 28 member states, and negotiating their immigration procedures can be difficult. Eliminating differences has been a core part of European Research Area policy, and in 2005 Europe-wide legislation was introduced to ensure fast-track immigration for international researchers through the 'scientific visa'. All EU countries except Ireland, the United Kingdom and Denmark were obliged to implement this fast-track visa, but the system has been put in place with varying levels of enthusiasm and effectiveness. Ireland recognized the potentially positive impact of the scientific visa and voluntarily introduced it in 2007 to attract more researchers.

Ireland's largely successful system can serve as a model for other nations. Although Ireland aspires to be a research hub, its economy remains fragile, and science funding is tight. Ireland cannot afford to stymie the movement of talent with cumbersome visa protocols.

We use a secure online verification system that links immigration services with embassies around the world, and we have a simple application procedure to reduce bureaucracy. The online system means that after applicants obtain research jobs in Ireland, they can get visas for themselves and their families within days, rather than the previous six to eight weeks for the scientists and up to a year for their families. Scientists used to have to pay €1,000 (US\$1,300) per year for a work permit; it is now free. Ireland implements the visa through its EURAXESS office — one of 200 offices across Europe providing advice and support for mobile scientists — based in Dublin, at the Irish Universities Association (of which I am the research director).

Over the past 6 years, 1,750 researchers from 80 countries have come to Ireland using the fast-track scientific visa. Half of them came



RUNNER: OORKA/SHUTTERSTOCK; FLAG: REDKOALA/SHUTTERSTOCK

from China, the United States and India. How do we know that the changes have made a real difference? In March, EURAXESS Ireland carried out a survey of more than 300 researchers with scientific visas, the first such survey by any European country participating in the scheme. Of those surveyed, 84% were still working in Ireland; 27% had found employment in the areas of information and communications technology, and 26% in computer and life sciences.

Fifty-three per cent of the respondents said that the fast-track visa was very important when deciding to proceed with a job in Ireland. The survey also revealed that 23% of researchers would definitely not have come to Ireland if the fast-track visa scheme had not been in place.

Policymakers and those keen on building a nation's scientific reputation often assume that facilities and staff are the crucial factors for drawing in researchers. However, our survey shows clearly that the immigration process is key to decision-making. The story is likely to be the same for other countries that are working to increase their scientific excellence by attracting researchers.

Competing for international scientific talent is a challenge. Nations such as Ireland that have invested heavily in science only in the past decade have to compete against international leaders including the United States and the United Kingdom. Providing an easily accessible, fast-track visa can make a real difference. ■

**Conor O'Carroll** is research director at the Irish Universities Association in Dublin and chair of the European Steering Group on Human Resources and Mobility.



# BOTTLED UP

*Hard to swallow.*

BY ANATOLY BELILOVSKY

“You won’t like it,” said the corkscrew. “A most disappointing product, not what I expect from a Grand Cru winery at all. Hardly any nose to speak of, and the tannins are totally unbalanced. You really should pull me out of this cork, lay the bottle down for a few more years, and make yourself a nice cup of tea.”

*He’d made it for her once and she’d pretended he’d got it right but he knew he hadn’t. The tea she made herself was so much better, its aroma suffused the whole house. She would sit in that enormous old chair in her comfy old sweater two sizes too big, feet tucked under her, nose in the teacup, eyes half shut behind steamed-over glasses, her smile would light up the room more than the fireplace ever did, and he would just sit and look at her for hours on end.*

*He might be looking at her still if she weren’t dead.*

“Just pop it,” he said through his teeth, eyes blurring, hands beginning to shake.

“Are you going to chug the wine?” asked the corkscrew.

He looked around, opened a cabinet, selected a glass.

“That’s a white-wine glass,” said the corkscrew.

“So?”

“The volatile scents won’t be properly concentrated,” said the corkscrew. “It will interfere with the proper enjoyment of the wine.”

“What are you, my mother?” he said.

“No, I am not your mother. I am Smork-screw, a beverage-management solution equipped with artificial intelligence. Now, what food were you considering to accompany this wine?”

He looked around again, pointed at a dish. “This?” he said.

“Peanuts?” said the corkscrew. “That would be an insult even to this inferior excuse for a Valpolicella. Haven’t you got any cheese? Fresh mozzarella would be perfect. Or turkey prosciutto and melon. Be sure to slice the prosciutto very thin —”

“She’s dead!” he shouted. “Can’t you get that through

your —” he waved his hands “— whatever you use instead of head!”

“I know she’s dead,” said the corkscrew. “It was my honour to open many a bottle for you and your guests as you sat *shiva*. Many people came.”

“Everyone loved her,” he said.

“She wasn’t there,” said the corkscrew. “She was dead, remember?”

“What’s your point?” he said. “Get it? Point?” He tried to laugh, but only a sob came out.

“They came for you,” said the corkscrew. “Many people. They all love you.”

“And now? Where are they now?”

“A phone call away.”

“Screw you,” he said. “Get it? Screw you?” He turned away, covered his face with his hands. “I want to be alone,” he whispered. “Just open the damn bottle.”

“She wouldn’t like it,” said the corkscrew.

“She is not *here* to not like it,” he bit off. “She is not in this room. She is not in this house. She isn’t anywhere on this Earth, anywhere under the Sun, anywhere except ...” He clenched his fists. “Open the bottle, please. Just ... I’ll get a proper glass, I’ll make some sandwiches, I’ll sit in her chair and cuddle her sweater, it won’t be like I’m drinking alone, I’ll drink to her and I’ll drink to me, just open it.”

“She’s not in that bottle, either,” said the corkscrew.

“What makes you such a damn expert on her?” he screamed.

“I am an expert system, you know,” said the corkscrew. “Able to access the totality of information about the two of you throughout the World Wide Web, and to correlate its implications upon the problem at hand.”

“Would my medical history be part of that totality?” he asked quietly.

“I must regretfully decline to divulge trade secrets of Smork-screw, LLC,” said the corkscrew. “However, consider this: I am — was — her present to you. The process of gift selection gave me an opportunity to perform a thorough assessment of her character and motivations.”

“What you mean is, she loved me, and she didn’t want me to drink myself to death,” he said. “In case ...”

He trailed off. An old clock ticked off the seconds in the silence. The cherry tree shook in the wind beyond the window, shedding the last of its dry brown leaves (*it had been white with bloom the day they packed to go on their second honeymoon, the sun had shone so bright, the shadows so crisp* —).

“She knew then, didn’t she?” he asked.

“Yes,” said the corkscrew.

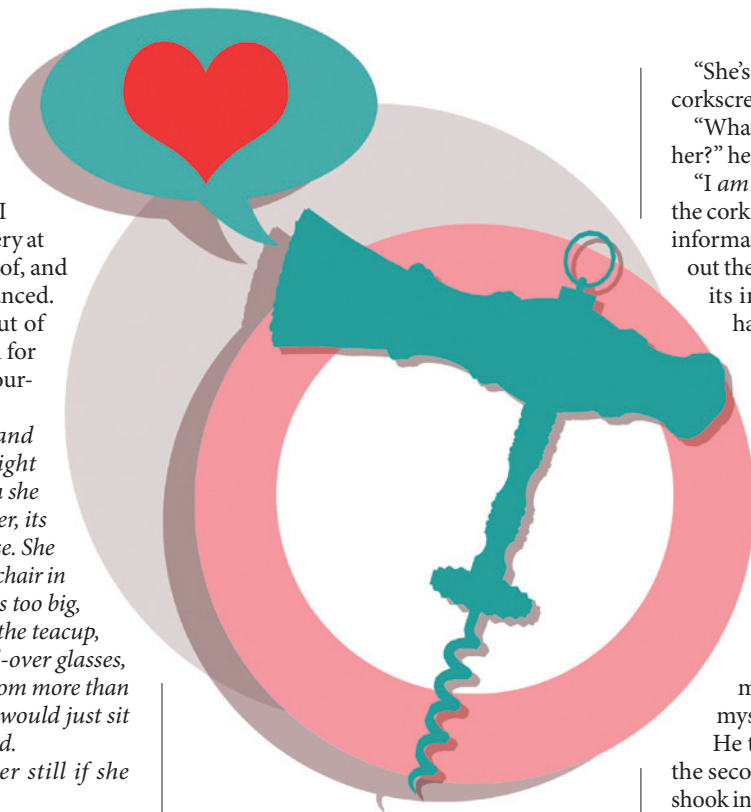
“She found out about the cancer, and then we went on the anniversary cruise, the best vacation ever, and she bought you for me as a present, and I thought it was cute and thanked her, but I had got her a pair of diamond earrings so it kind of irked me for a while that all she got me was a stupid corkscrew and all I said was, ‘Cute,’ and then a week after we got back she told me she was going to die, and then she died and I never said I was sorry ...”

“Apology accepted,” said the corkscrew.

It wound itself from the cork and rolled towards its drawer, pausing at the counter’s edge as he shook with sobs too long in coming. ■

Anatoly Belilovsky was born in what is now Ukraine, learned English from Star Trek reruns, worked his way through a US college by teaching Russian while majoring in chemistry, and has, for the past 25 years, been a paediatrician in New York, in a practice where English is the fourth most commonly spoken language. He blogs at <http://loldoc.net>.

JACEY



➔ NATURE.COM

Follow Futures:

@NatureFutures

go.nature.com/mtoodm